Proteomic studies of the hid1Δ and hid3Δ mutants of Schizosaccharomyces pombe

Abdulrahman Alasmari

Année 2015

Thèse n°2015BORD0142

THÈSE

pour le

DOCTORAT DE L'UNIVERSITÉ de BORDEAUX

Ecole Doctorale des Sciences de la Vie et de la Santé

Spécialité: Génétique

Présentée et soutenue publiquement

Le 16 Septembre 2015

Abdulrahman Alasmari

Né(e) le18 Décembre 1984 à Riyadh

Proteomic studies of the *hid1Δ* and *hid3Δ* mutants of *Schizosaccharomyces pombe*

Membres du Jury

M. Moreau, Patrick	Univeristé de Bordeaux	Président
M. Gachet, Yannicl	Université Paul Sabatier	rapporteur
M. McFarlane, Ramsay	Bangor University	rapporteur
M. El Ghouzzi, Vincent	INSERM, Hôpital Robert Debré	Examinateur
M. Hooks, Mark A.	Univeristé de Bordeaux	Directeur de thèse

Études protéomiques des mutants *hid1Δ* et *hid3Δ* chez *Schizosaccharmyces pombe*

RÉCAPITULATIF EN FRANÇAIS

Schizosaccharomyces pombe est devenu un important système modèle pour étudier les processus physiologiques, biochimiques et génétiques chez l'homme. Ces travaux appartiennent à un projet sur la façon dont la fonction altérée de l'appareil de Golgi contribue à des maladies comme le cancer ou cause des anormalités génétiques. La protéine HID-1 de *C. elegans* et des humains est une protéine de l'appareil de Golgi qui appartient à la superfamille de protéines DYMECLIN. Les animaux ont à la fois un gène *HID1* et un gène *DYM.* Chez les humains, une expression réduite de *HID1* est impliquée dans la prolifération de tumeurs. La perte de DYM chez les humains mène à une déformation squelettique. *S. pombe* a trois gènes orthologues HID-1, mais pas de DYM. Par contraste, de nombreux eucaryotes unicellulaires et pluricellulaires n'ont qu'un gène DYM. Les mutants de *S. pombe* sans Hid1 et Hid3 étaient sensibles au stress oxydatif et la croissance de *hid3Δ* a été stoppée dans des milieux de culture minimum standard. L'insensibilité de *hid3Δ* au brefeldin A, mais sa sensibilité au golgicide, a démontrent que Hid3 fonctionne dans le transport antérograde à travers l'appareil de Golgi. Afin d'explorer des rapports indiquant que le renouvellement des protéines pourrait être modifié dans *hid3Δ*, j'ai entrepris une étude de protéomique a la quantification label-free. La régulation positive de la voie de signalisation MAPK de tension a démontré que les cellules étaient dans un état de tension dans des conditions normales de croissance. De plus, des composants de protéine dans plusieurs voies de signalisation étaient modifiés, pouvant affecter une large gamme de processus cellulaires.

MOTS CLÉS: Formation Dauer, Proteomique, Signalisation de stress

ENGLISH SUMMARY

Schizosaccharomyces pombe has become an important model system to study physiological, biochemical and genetic processes in humans. This work is part of a continuing project to study how altered Golgi function contributes to diseases, such as cancer, or causes of genetic abnormalities. The HID-1 protein of *C. elegans* and humans are peripheral membrane proteins of the Golgi apparatus and are part of the DYMECLIN superfamily of proteins. Animals have both a *HID1* and a *DYM* gene. In *C. elegans*, HID-1 maintains normal cellular growth and in humans reduced expression of *HID1* has been implicated in tumour proliferation. Loss of DYM in humans leads to skeletal deformation and potentially mental retardation. *S. pombe* has three HID-1 orthologues, but no DYM. In contrast, many unicellular and multicellular eukaryotes have only DYM. *S. pombe* mutants lacking Hid1 and Hid3 were sensitive to oxidative stress and growth of *hid3Δ* was stopped in standard minimal media. Insensitivity of *hid3Δ* to brefeldin A but sensitivity to golgicide A demonstrated that Hid3 operates in anterograde protein transport through the Golgi. In order to investigate reports that protein turnover might be altered in *hid3Δ*, I undertook a proteomics study using label-free protein quantification. Up-regulation of the MAPK stress signalling pathway demonstrated that cells were under a state of stress under standard growth conditions. In addition, protein components of Ras signalling, microtubule dynamics and chromatin remodelling were altered potentially affecting a wide variety of processes from cell cycle regulation to metabolism.

KEY WORDS: Dauer Formation; Proteomics; Stress Signaling

Table of Contents

List of Tables and Figure

Acknowledgments

This work was funded by Tabuk University in Saudi Arabia. I extend my fondest thanks to my family, who supported me during my PhD journey. My deep gratitude also goes to my supervisor, Dr. Mark Hooks, for all his guidance, support and encouragement. Many thanks are extended to all the people who supplied me with strains, especially Dr. David Pryce. I also would like to thank Dr. Marc Bonneu and Dr. Stephane Claverol of the Proteomics Platform at the Centre Génomique Fonctionelle de Bordeaux for conduction the protein quantification proteomics data. I also gratefully acknowledge the support from all my colleagues, past and present, and especially the members of the INRA Labs.

Abbreviations

1-D:	one-dimensional gel electrophoresis
^{1}H NMR :	H-NMRhydrogen-1 NMR
AtDYM:	DYM from Arabidopsis
ATM:	ataxia telangiectasia mutated
Aβ:	amyloid-beta
BFA:	Brefeldin A
Bp:	base pairs
BSA	Bovine Serum Albumin
C.elegance:	Caenorhabditis elegans
c10orf76:	chromosome 10 open reading frame 76
c17orf28:	chromosome 17 open reading frame 28
CDKs:	catalytic cyclin-dependent kinases
CeDYM:	DYM from C. elegans
CESR:	core environmental stress response
CMM:	Cisternal Maturation Model
CPT:	camptothecin
CSDE1:	cold-shock domain-containing protein
Da:	Dalton
Daf-c:	Dauer formation consistitutive
Daf-d :	Dauer formation defective
DMC1:	DOWNREGULATED IN MULTIPLE CANCERS 1
DNA:	deoxyribonucleic acid
DNA-PK:	DNA-dependent protein kinase
DR:	down-regulated or less abundant
DUB:	deubiquitinating enzyme
DYM:	Dyggve-Melchior-Clausen syndrome (DYMECLIN)
e.g.	for example
EDTA:	Ethylenediaminetetraacetic acid solution
EMM:	Edinburgh Minimal Medium
ER:	Endoplasmic Reticulum
ERAD:	Endoplasmic reticulum-associated degradation

ERES. Endoplasmic reticulum exit sites
ERK: extracellular-signal regulated kinase
G418: Geneticin
GCA: Golgicide A
GE: Generation Time
GEFs: guanine nucleotide exchange factors
GO analysis: Gene Ontology Analysis
GRASPs: Golgi reassembly stacking proteins
HID-1: Heat-Induced Dauer formation protein 1
HIS: histidine
HR: homologous recombination
HsDYM: DYM from humans
HSPs: heat shock proteins
HU: hydroxyurea
i.e. that is
ITRAQ: Isobaric tag for relative and absolute quantitation
JAMM : JAB1/MPN/Mov34 metalloenzyme domain proteases
KCl: Potassium chloride
L1-L4: four larval stages
LC-MS/MS: liquid chromatography-tandem mass spectrometry
LOH: Loss of Heterozygosity
Lycopsida: lycophyte
m/z: mass to charge
MALDI-ToF MS: ionization time of flight mass spectrometry
MAPK: Mitogen-activated protein kinase
Min: minutes
MJD : Machado-Josephin domain proteases
MMS: methyl methanesulphonate
mRNA: Messenger RNA
MS/MS: tandem mass spectrometry
MS: mass spectrometry
MTC: mitomycin C
Multi-Ub: multi ubiquitin
Nat: nourseothricin-resistance gene

NCBI: National Center for Biotechnology Information
NMR: nuclear magnetic resonance
ORFs: open reading frames
OTU : ovarian tumour proteases
PAGE: polyacrylamide gel electrophoresis
PCA : Principal Component Analysis
PCR: Polymerase Chain Reaction
PI3K: phosphoinositol-3-Kase
Poly-Ub: poly ubiquitin
P. pastoris: Pichia pastoris
PVDF: Polyvinylidene fluoride membrane
r.p.m.: revolutions per minute
RNA: ribonucleic acid
ROS: reactive oxygen species
S. cerevisiae: Saccharomyces cerevisiae
S. pombe: Schizosaccharomyces pombe
SAPK1: stress activated protein kinase pathway 1
SDS: sodium dodecyl sulfate
SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Sec: seconds
SESR: stress-specific environmental stress response
SMC: McCourt dysplasia
SREBP: Sterol Receptor Element Binding Protein
SUMO: small Ub-like modifier
TBZ: thiabendazole
TGN: trans-Golgi network
TOF: time of flight
TSGs: Tumour Suppressor Genes
Ub: ubiquitin
UBA: ubiquitin associated domain
UBL: ubiquitin-like domain
Ubp: ubiquitin specific protease
UCH : ubiquitin C-terminal hydrolases
UR: Up-regulated

URA: uracil

USPs : ubiquitin-specific proteases

UV: ultra violet

URA: uracil

USPs : ubiquitin-specific proteases

UV: ultra violet

VCN: Vector controls Nat^R

WT: wild type

YE: yeast extract

YES: Yeast extract with Supplements

Récapitulatif du projet en Français

Introduction

Les protéines destinées à être sécrétées ou transférées vers un compartiment particulier sont fabriquées dans le réticulum endoplasmique (RE), puis elles doivent être traitées pour les rendre entièrement fonctionnelles. C'est une des principales fonctions de l'appareil de Golgi. L'appareil de Golgi est divisé en compartiments appelés citernes. Les protéines issues du RE pénètrent dans l'appareil de Golgi par la face-*cis* et ressortent par la face-*trans.* Il en résulte que toutes les protéines doivent passer dans la région dite de citerne, qui contient les enzymes divers, appelés enzymes résidents, pour faire les modifications comme la glycosylation, la sulfatation ou la phosphorylisation.

L'Ubiquitine (Ub) est une petite protéine (76 aminoacides), compacte et superpliée, qui est fortement conservée dans les organismes eucaryotes. Il existe trois différences d'acides aminés entre la levure et l'homme (Shaid et al., 2013). Les cellules utilisent Ub comme une étiquette pour modifier le contenu des protéines en fonction des conditions cellulaires pendant la dégradation des protéines. La voie Ub est aussi une modification post-translationnelle des protéines qui est importante et étendue, mais ne se limite pas à la dégradation. L'ubiquitylation guide la fonction protéique dans la réparation de l'ADN, le transfert protéique, la modification de la structure protéique, la modulation de l'activité des protéines cibles, la croissance cellulaire et la transduction de signal (Horák, 2003). L'ubiquitination est un processus important dans le système de la membrane endosomale qui comprend le réticulum endoplasmique et l'appareil de Golgi. L'ubiquitylation est importante et permet aux cellules de vérifier que seules les protéines fonctionnelles sont modifiées et finalement sécrétées. La plupart des protéines qui sont ubiquitinées dans une cellule sont des protéines nouvellement synthétisées qui sortent juste de ribosomes (Kim et al., 2011) et il est estimé que l'ubiquitylation pourrait servir de système de contrôle de la qualité pour les protéines nouvellement synthétisées (Chhangani et al., 2012). Cela constitue aussi une partie d'un processus important de signalisation pour permettre à l'appareil de Golgi de communiquer avec le noyau. Par exemple, le contrôle des niveaux de cholestérol dans les cellules humaines passe par la voie de la protéine se liant à l'élément récepteur du stérol (SREBP). Autant chez l'homme que *S. pombe*, le clivage de l'extrémité N-terminale de SREBP dans la membrane de l'appareil de Golgi est initié par son ubiquitylation (Stewart et al., 2011). Au cours des dernières années, un grand nombre de maladies a été lié

au mauvais fonctionnement de l'appareil de Golgi, à l'inclusion de maladies cutanées, de troubles neurologiques, de dysplasies squelettiques et de troubles des tissus conjonctifs (Bexiga and Simpson, 2013).

Le gène de la *protéine 1 de formation de larves dauer induite par des températures élevées* (*hid-1*) de *C. élégans* a été trouvé dans un dépistage à la recherche de mutants du nématode, qui a adopté la formation de larves Dauer dans de bonnes conditions de croissance, mais avec une température légèrement plus élevée (Ailion and Thomas, 2003). HID-1 est une protéine de la membrane de l'appareil de Golgi qui semblerait être impliquée dans l'excrétion de neuropeptides dans *C. elegans.* Il a été montré que le gène orthologue *S. pombe* de HID-1 localise le déubiquitinase Ubp5 (gène orthologue S. pombe de USP7 chez l'homme) dans l'appareil de Golgi. La translation fonctionnelle chez l'homme est qu'un mauvais fonctionnement de la protéine HID1 ou son absence affectera la place dans la sous-cellule de l'enzyme déubiquitylatant USP7.

Objectifs du projet et organisation de la thèse

Le but de l'introduction fournie dans le chapitre 1 était de présenter l'idée que la modification de la fonction de l'appareil de Golgi a des effets au delà du transport vésiculaire. Ceci est intimement lié au fonctionnement général des cellules à travers un certain nombre de systèmes de signalisation. L'ubiquitinylation était au centre de cette introduction parce que la manipulation des processus Ub est une des fonctions proposées de la protéine Ftp105/Hid3 de *S. pombe.* Toutefois, il est probable que la protéine Ftp105/Hid3 ait d'autres fonctions dans l'appareil de Golgi ou qu'elle puisse affecter d'autres processus quand elle est mutante. La question demeure aussi concernant le rôle des deux gènes paralogues Ftp105/Hid3. Les travaux du groupe, dont ce projet de thèse fait partie, utilisent des technologies postgénomiques de transcriptomique, protéomique et métabolomique pour avoir une meilleure idée de ce que font ces protéines, pour étudier les réactions des mutants à la tension sans remplacer les gènes *hid*, puis pour déterminer les conséquences sur le métabolome et le protéome des mutants. Les données de ces travaux sont présentées dans trois parties principales (chapitres 3 à 5) pour finir par une discussion générale sur le projet et une proposition d'orientations futures des travaux (chapitre 6). Les résultats sont résumés pour chacune de ces parties. Le chapitre 2 détaille la méthodologie et n'a donc pas été inclus.

Chapitre 3: Analyses de calcul de la superfamille de protéines DYMECLIN et les gènes orthologues HID-1 issus de *S. pombe*

Le chapitre 3 décrit les travaux de calcul pour analyser les HID et leurs relations avec leur membre le plus proche, le DYMECLIN (DYM). Les travaux phylogénétiques ont été conduits pour déterminer s'il pouvait être prouvé que des fonctions des deux protéines se chevauchaient. La présence ou l'absence de HID et DYM a été vérifiée dans les supergroupes et les taxons de premier rang de NCBI (tableau 3.1). Les animaux présentaient des protéines HID et DYM, leur lieu et leur action, comme l'association dynamique avec l'appareil de Golgi, semblaient être très similaires. Il est évident que les eucaryotes peuvent avoir des protéines DYM et/ou HID tandis que les animaux ont les deux. Les champignons, sauf quelques exceptions rares, n'ont que des protéines HID et les plantes n'ont que des DYM (figure 3.1, 3.2). La phylogénie et les comparaisons de séquences ont confirmé la présence de trois gènes orthologues HID1 dans *S. pombe* et présentaient une séquence de diversification (figure 3.3, tableau 3.2). J'ai proposé que Ftp105 soit appelé Hid3 à l'avenir. *S. pombe* représente un bon modèle pour étudier la fonction de Hid, parce que cela élimine les complications potentielles de la présence de protéines DYM, mais il est possible qu'il y ait une fonction redondante au sein des SpHid.

Des informations de bases de données disponibles au public indiquent que Hid1 et Hid2 se trouvent dans le cytosol/noyau avec une fonction inconnue, tandis que Hid3 est une protéine de l'appareil de Golgi (mais aussi peut-être cytosolique) et responsable de la position des protéines dans l'appareil de Golgi (tableau 3.3). Une analyse structurelle des données de séquence suggèrent que SpHid1 est plus étroitement liée aux HID animaux, mais il semble que SpHid3 a conservé les propriétés structurelles qui lui permettent de se fixer à l'appareil de Golgi (figure 3.4). Enfin, des données disponibles au public ont été examinées pour découvrir des associations possibles de chacun des Hid avec d'autres protéines qui pourraient révéler des aspects de fonction (figure 3.5). Les Hid n'ont présenté qu'un nombre très faible d'associations, mais il a été démontré que Hid3 était associé à Ubp5 comme des rapports l'ont indiqué. De cette association, un modèle de base a été créé pour montrer comment la modification des niveaux de protéine Hid3 peut affecter les niveaux d'autres protéines à travers Ubp5 et au final le métabolisme et l'expression génique (figure 3.6).

Chapitre 4 : Effet des traitements de la tension sur la croissance de mutants *hidΔ*

Des données déjà publiées suggéraient que *hid1*$^{+}$ et *hid2*$^{+}$ peuvent être impliquées

dans les réactions à la tension, parce qu'elles sont induites en réaction à une tension (figure 4.1). Il a été rapporté que le mutant *hid3* est sensible à certains traitements, comme des doses de sel élevées et certains produits pharmaceutiques (www.pombase.org). Les mutants produits dans l'étude d'Alshehri (2015) ont donc été testés pour déterminer leurs réactions de croissance à des tensions/traitements chimiques divers (tableaux 4.1, 4.2). Aucun des mutants simples *hid1Δ*, *hid2Δ* et *hid3Δ* n'était sensible à une gamme variée de tensions, sauf au stress oxydatif, qui affectait *hid1Δ* et *hid3Δ* en ralentissant fortement la croissance des cultures (figure 4.2). Il est intéressant de noter que la croissance de *hid3Δ* a été sérieusement affectée quand elle était cultivée dans des milieux de culture minimum (figure 4.3). On n'en connaît pas la cause, mais je spécule que *hid3Δ* manque d'un nutriment essentiel. Pour finir, en utilisant les inhibiteurs de l'appareil de Golgi brefeldin A (figure 4.4) et golgicide A (figure 4.5), j'ai pu déterminer que dans *S. pombe*, Hid3 fonctionne dans le transport antérograde des protéines (figure 4.6).

Chapitre 5 : Caractérisation métabolomique et protéomique des mutants *hid1Δ* et *hid3Δ* de *S. pombe*

Ce chapitre décrit les propriétés métabolomiques et protéomiques des mutants. La métabolomique sert à juger dans quelle mesure la physiologie et la biochimie fondamentales des cellules se modifient quand une mutation génétique est introduite. Par ailleurs, les évolutions métaboliques peuvent fréquemment expliquer pourquoi la croissance cellulaire est modifiée. Les échantillons utilisés dans l'étude étaient exactement les mêmes que ceux utilisés dans l'étude transcriptomique (tableau 5.1). Un nombre limité de métabolites cellulaires a été quantitativement mesuré avec ^{1}H-NMR. Ces mesures étaient particulièrement remarquables en raison de l'augmentation de la quantité de glycérol et de tréhalose présents dans *hid3Δ*. Ces composés sont synthétisés en réaction à la tension, mais ils sont aussi produits en adaptant les cellules à un état quiescent. Une analyse en composantes principales des données du métabolite a montré que l'état métabolique de *hid1Δ* et de *hid3Δ* étaient différents de la variation contingente négative (VCN) de contrôle (figure 5.1).

La conduite d'une étude protéomique visait un but directement lié à l'implication potentielle de Hid3 dans les processus utilisant l'ubiquitine qui impliqueraient la dégradation des protéines. Il est clair que le champ de la protéomique est large et qu'il peut y avoir plusieurs façons dont l'affectation de l'appareil de Golgi peut modifier les niveaux protéiques, mais la première étape était de déterminer les protéines qui évoluaient et de combien. Des protéines totales ont été extraites des mêmes échantillons utilisés pour l'étude métabolomique. Par la

suite, une électrophorèse SDS-PAGE a montré que les génotypes ne présentaient pas de grandes différences au niveau de l'abondance des protéines, puisque les schémas présentés par les bandes et les quantités de protéines apparaissaient uniformes (Figure 5.2). Les échantillons de protéine ont été envoyés à l'unité de protéomique de CGFB où 10 µg de la protéine totale de chaque échantillon a été purifiée par gel et digérée dans le gel avec de la trypsine (figure 5.3). Les protéines ont été quantifiées par une analyse du mélange résultant de peptides au moyen de la spectrométrie de masse (tableau 5.3). L'analyse en composantes principales (figure 5.4) et la quantification de protéines individuelles (Figure 5.5) a montré que la composition de la protéine *hid3Δ* était fortement différente de celle d'autres génotypes. L'analyse ontologique des protéines en évolution a montré pour *hid3Δ* des niveaux élevés de protéines de réaction à la tension et des niveaux réduits de protéines biosynthétiques (figure 5.6), indiquant des cellules dans des conditions chroniques de tension. L'inspection de protéines individuelles qui ont été modifiées en abondance a révélé la régulation positive dans *hid3Δ* de points clés de la voie de signalisation MAPK de réaction à la tension (Tableau 5.5). D'autres protéines intéressantes faisant l'objet d'une augmentation étaient les protéines de remodelage de la chromatine Obr1 et l'histone H2B qui affecteraient l'expression génique et la GTPase Ras1, qui est une protéine clé de voie de signalisation dont la fonction dépend de sa place au sein de la cellule. Ces différences montrent que l'élimination de Hid3 engendre un éventail large d'effets sur la fonction cellulaire.

Chapitre 6 : Discussion générale et travaux futurs

Ce chapitre présentait les conclusions obtenues par chaque partie de l'étude. Les principales conclusions du chapitre 3 étaient la confirmation de plusieurs gènes orthologues HID1 dans *S. pombe* et l'absence de DYM. En conséquent, *S. pombe* est un modèle qui convient dans l'étude de la fonction de Hid. Du fait de leur absence de HID, d'autres organismes comme les plantes plus grosses pourraient être utilisés pour étudier la fonction de DYM. L'étude phylogénétique n'a pas abouti à la fondation de l'origine de la superfamille de DYM, mais il est clair que ces protéines sont apparues avec la diversification majeure des organismes eucaryotes. La caractérisation phylogénétique de la superfamille de DYM demande des travaux plus importants. Les principales conclusions des tests de tension présentées au chapitre 4 étaient la sensibilité de *hid1Δ* et *hid3Δ* aux traitements oxydants, l'absence de croissance de *hid3Δ* dans des milieux de culture minimum et la confirmation du fonctionnement de Hid3 dans le mouvement antérograde des protéines. Il n'est pas clair pourquoi *hid1Δ* a été affecté par le

traitement à l'eau oxygénée pure et pourquoi le mutant *hid2Δ* pouvait inverser l'effet. Il est probable que *hid3Δ* subissait déjà des tensions et que des réactions étaient déjà activées de sorte qu'il ne pouvait pas gérer l'insulte supplémentaire du traitement à l'eau oxygénée pure. L'incapacité de *hid3Δ* à se développer dans des milieux de culture minimum suggère qu'il lui manque un nutriment essentiel qui peut être acquis par des souches non affectées. Ceci va dans le même sens que l'induction de l'expression des gènes encodant les protéines de transport (Alshehri 2015). Le chapitre 5 a confirmé autant pour le métabolisme que les niveaux de protéines que les cellules de *hid3Δ* subissent une forte tension même dans des conditions normales de croissance. Les évolutions des niveaux de métabolite et de protéine sont cohérentes avec l'idée que la croissance réduite des cellules de *hid3Δ* peuvent être dans un état de quiescence partielle et les évolutions des niveaux de protéine présentent le mécanisme qui explique les modifications observées dans l'expression génique (figure 6.1). Toutefois la gamme large de protéines dont les niveaux ont été modifiés montre que de nombreux processus différents seront affectés dans les souches de *hid3Δ.* Enfin, je discute de plusieurs voies possibles de recherche allant de l'utilisation des HID et DYM humains et de *C. elegans* pour inverser les phénotypes *hid3Δ* jusqu'à la caractérisation biochimique et structurelle des associations de Hid3 avec la membrane de l'appareil de Golgi. Il faudra déterminer l'emplacement dans les sous-cellules des protéines de signalisation, comme Ras1 ainsi que les mécanismes qui servent à modifier les protéines soit par l'augmentation de l'expression des gènes, soit par la modification de la dégradation.

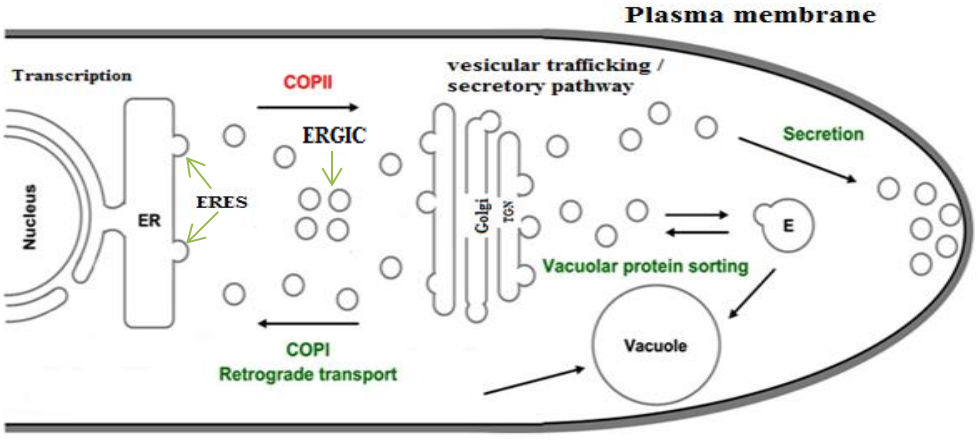


Figure 1.1. Trafficking through the endomembrane system in yeast. Proteins synthesized in the rough endoplasmic reticulum (ER) are passed to the Golgi apparatus by COPII protein complex coated vesicles in endoplasmic reticulum (ER)-Golgi intermediate compartment (ERGIC). As the proteins and lipids traverse the Golgi, they are processed and/or modified and subsequently sorted to various compartments within the cell, such as the vacuole of plasma membrane, or secreted outside of the cell. The image was taken from Aguiar et al., (2014).

Chapter 1

General Introduction

1.1 Function of the Golgi apparatus in the cell

There are two main types of cells, prokaryotic and eukaryotic, which form the basis for cellular life. They carry the basic things that all cells require, such as a cell membrane, DNA, cytoplasm, and ribosomes (Alberts et al., 2002). Eukaryotic cells differ from prokaryotic cells in that they can be unicellular like prokaryotes but have a diverse range of cell shapes and sizes and can combine to form multicellular organisms. On the cellular level there are fundamental differences between eukaryotic and prokaryotic cells, such as a nucleus, and organelles, such as mitochondria or plastids, that can generate energy. Eukaryotic cells also have a well-defined and intricate endomembrane system that transports proteins to different parts of the cells and allows cell to cell communication in multicellular eukaryotes. The group is interested in a part of this endomembrane system known as the Golgi apparatus, which is responsible for the processing and packaging for sorting proteins and lipids to other parts of the cell, such as the lysosome in animal cells or the vacuole in plants and fungi, or to the outside of the cell (Glick, 2000). The Golgi is an organelle that consists of various numbers of membrane stacks that various in number and shape depending on the organism. The Golgi apparatus is comprised of four main parts, *cis*-Golgi, *medial*-, *trans*-Golgi and trans-Golgi network (TGN). The terms '*cis*' and '*trans*' refer to the different faces of a Golgi complex. Vesicles arrive at the cis-face from the Endoplasmic Reticulum (ER) and they leave from the trans-face to travel to the plasma membrane or an interior compartment (Figure. 1.1) (Nakano and Luini, 2010).

Proteins are found everywhere in the cell as they are the 'workers' of the cell providing structure and catalysing reactions. Protein function is so closely related that when a protein does not function properly it can have bad effects on a cell leading to different types of congenital diseases, such as muscular dystrophy, Alzheimer's disease and cystic fibrosis. Most importantly for my study is the group of proteins that can be directed to specific places in the cell, for instance the plasma membrane, the cytoplasm, the nucleus, the endosomes/vacuole, and the cellular machinery that accomplishes this, namely the Golgi apparatus. The diseases mentioned above relate to disrupted protein trafficking by Golgi apparatus (Ungar, 2009).

Proteins destined for secretion or for transfer to a particular compartment are made in the ER and then they need to be processed to make them fully functional. That is one of the

main functions of the Golgi apparatus. The Golgi is divided into compartments called cisternae. Proteins enter the Golgi from the ER by *cis*-face and they exit from *trans*-face side.

As a result, all proteins have to pass a region called cisterna which contains the various enzymes, called resident enzymes, for doing protein modifications, such as glycosylation, sulfation, or phosphorylation. There are two theories about how proteins move through the Golgi, the Cisternal Maturation Model (CMM) theory and the Vesicular Model (VM) theory. In the CMM, the Golgi cisternae themselves mature and convert to the next type and in doing so carry their cargo from *cis*-face to *trans*-face. The VM model states that proteins are passed to fixed cisternae through vesiculation (Connerly, 2010). Using GFP-labelled proteins in budding yeast, two independent laboratories provided evidence for the CMM model to show that resident proteins move through the stacks (Losev et al., 2006; Matsuura-Tokita et al., 2006), so this remains the preferred theory. However, both types of movement are likely to occur, because proteins can be recycled back through the Golgi, which is likely mediated by vesicle transfer (Emr et al., 2009).

1.1.1 Golgi in *S. pombe*

There are a number of different Golgi structures that have been identified in various organisms, but the typical structure as *cis*-, *medial*-, and *trans*-cisternae plus the trans-Golgi network (TGN). These compartments have been found to differ from each other in properties, like cytochemical-staining, composition of resident enzymes, and ability to bud COPI- or clathrin-coated vesicles (Farquhar and Palade, 1981; Kleene and Berger, 1993; Nilsson et al., 2009; Rabouille et al., 1995; Staehelin and Kang, 2008). Resident Golgi proteins can move rapidly within and between compartments. For example, according to the cisternal maturation model mentioned above, glycosylation enzymes that operate in assembly line fashion to process secretory cargoes must be re-transported by to the original site of function if not degraded (Glick and Luini, 2011; Nilsson et al., 2009; Rabouille et al., 1995). A study by Glick and Luini, (2011) tried to investigate various models known to be linked to Golgi traffic, such as vesicular transport between compartments and cisternal maturation, to provide a unifying observation for Golgi function. They concluded that no single model can easily explain all of the observations from diverse organisms. A study by Klute et al., (2011) compared the origin and history of the Golgi by comparing proteins involved in Golgi morphology as well as those involved in trafficking, by choosing proteins that act as carriers to and from the Golgi at both the *cis*- and *trans*-faces. They concluded that the basic elements of vesicle trafficking and Golgi structural appear conserved across eukaryotes.

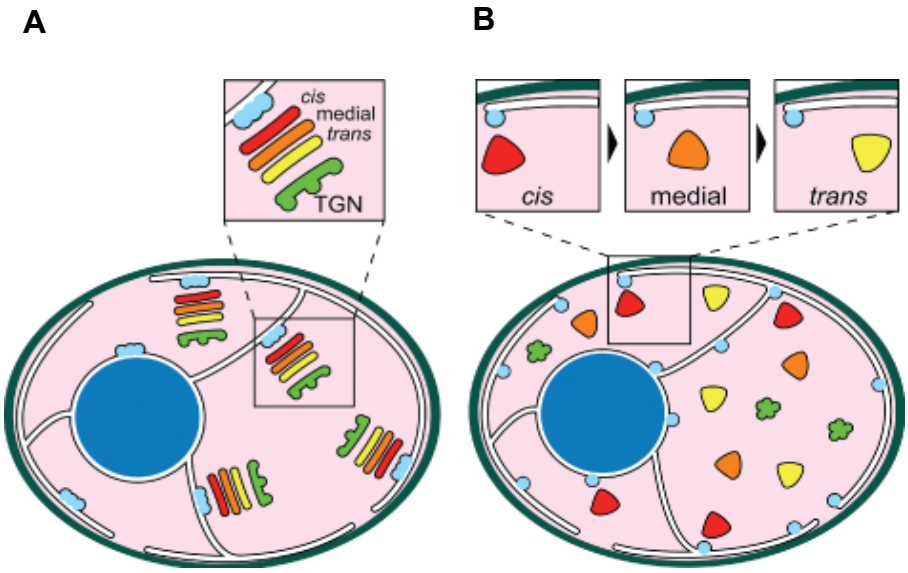


Figure (1.2). Diagram of the Golgi apparatus and ER exit sites in yeast species. **(A)** Ordered Golgi stacks locate close to the ERES (light blue) in *P. pastoris* similar to *S. pombe*. *The cis* cisterna (red) of Golgi stack faces ERES. **(B)** in *S. cerevisiae*, Golgi cisternae are it matures into a medial and a *trans* cisterna and dispersed to the cytoplasm (Suda and Nakano, 2012).

Many of the processes of the Golgi complex of higher eukaryotes like mammals, such as carbohydrate modification and proteolytic processing, have been found in to occur in the yeast Golgi (Seemann et al., 2000).There would not be expected to be identical functions of the Golgi between higher eukaryotes and the yeast, because the Golgi in higher eukaryotes is necessary for processing and secreting many of the factors that are necessary growth, proliferation and cell communication within a multicellular context. The internal cell structure of mammalian cells is also different to that of yeast, where mammalian cells have an endosomal system for degrading cellular components, where yeasts have a vacuole system. The structure of the Golgi apparatus among yeast is also different where yeasts like *S. pombe* (Chappell and Warren, 1989) and *P. pastoris* (Preuss et al., 1992; Rossanese et al., 2001) have a stacked lamellae structure like higher eukaryotes, but *S. cerevisiae* has a more dispersed type of Golgi structure lacking visible cisternae (Figure 1.2). However, Mammalian-like Golgi reassembly stacking proteins (GRASPs) have been discovered in the three yeast species noted as factors involved in the formation of Golgi stacks (Vinke et al., 2011). Moreover, in yeast and unlike the mammalian cells, the Golgi down during the cell division to partition during cytokinesis (Preuss et al., 1992; Rossanese et al., 2001).

However, even in a dispersed state, the Golgi must be functional in *S. pombe* in order to form the septum. More investigations are needed to identify the similarities and differences in the nature of the Golgi complex, because the Golgi is an ancient and ubiquitous feature of eukaryotic cells. In addition, the use of various yeast species will help us to understand the importance of Golgi stack formation.

In addition to the morphology of the Golgi, there are functional differences among the yeasts as revealed in differences in protein components. For example, Rho3 has distinct roles in *S. pombe* and *S. cerevisiae*. In *S. cerevisiae* it appears to be involved in influencing cell growth by regulating polarized secretion through the actin cytoskeleton by interacting with Exo70 and Myo2 (Robinson et al., 1999). In *S. pombe*, Rho3 plays an important role in Golgi/endosome trafficking by complexing with Apm1 and other subunits of the AP-1 complex. Another main feature differing *S. pombe* from *S. cerevisiae* and other eukaryotes is that it that has only one Ras protein. There are two Ras proteins in *S cerevisiae* and four in mammals. This *S. pombe* Ras displays differing localisation (Fukui et al., 1986; Nielsen et al., 1992), which is known to influence the pathways it controls, such as the byr2/byr1/spk1 mitogen- activated protein kinase cascade. It also activates the the Rho GTPase Cdc42 protein (García et al., 2006) pathway to regulate cell polarity by modulating cytoskeletal elements and maintaining elongated cell morphology and protein trafficking. The differential location of Ras raises important questions as

to how do signalling proteins selectively regulate particular pathways. Mammalian and *S. pombe* Ras proteins have significant sequence homology that makes *S. pombe* an ideal organism to study aspects of Ras signalling, as it contains only a single Ras protein.

1.2. The role of ubiquitylation

Ubiquitin (Ub) is a small (76 amino acids), compact and super-folded protein which is highly conserved in eukaryotic organisms. Between yeast and humans there are only three amino acid differences (Shaid et al., 2013). Cells use Ub as a tag to change according to cellular conditions. The cell must respond to changing conditions by producing the proteins it requires at the moment. Accordingly, it is beneficial to the cell to degrade or recycle those proteins that are not required. Ubiquitylation has evolved to be a principal mechanism where proteins are tagged by direct binding with Ub to tag them for degradation by the proteasome (Hershko and Ciechanover, 1998). The Ub/proteasome pathway also is important to keep the cells healthy and functioning in presence of translation errors and genetic mutations that damage proteins (Tu et al., 2012). Misfolded proteins can be rescued by repair proteins called chaperones, but if not, the protein will be directed to degradation by the Ub/proteasome pathway. The Ub pathway is an important and widespread post-translational modification of proteins, but not only for degradation. Ubiquitylation guides protein function in DNA repair, protein translocation, altering protein structure, modulating activity of the target proteins, cell growth and signal transduction (Horák, 2003).

There are also different kinds of Ub-like proteins (UBLs), such as SUMO (small Ub-like modifier), which have some similarities in structure with Ub. In humans, there are 10 genes encoding UBLs called NEDD1 – NEDD10. Like Ub, NEDD UBLs are conserved among eukaryotes. They also have the ability to attach the target proteins to modify their function or stability, and have an important role in cell cycle control and embryogenesis (Brown and Jackson, 2015). The process of NEDD-UBL ligation is known as neddylation and it shows similarity to ubiquitylation in that NEDDs must be processed before attachment to protein substrates.

1.2.1 Ub ligases and deubiquitinating enzymes

Ub is added covalently to a target lysine residue on a protein target either as a single protein, as a chain of Ub proteins (poly-Ub) and/or on multiple sites on the same protein (multi-

Ub). Poly-Ub is the process for tagging proteins for degradation, whereas mono-Ub generally affect protein localization and multi-Ub can affect protein function (Xu and Jaffrey, 2011). The process of ubiquitylation is a three-step process performed by what are called Ub ligases. E1 ligase is the first step and activates Ub for attachment to Ub E2 ligase, which is also called ubiquitin conjugating enzyme. Thirdly, Ub is transferred from E2 onto the C-terminus of E3, which then attaches Ub on to specific lysine residues of target proteins by ac C-terminal glycine. E3 Ub ligases are the most abundant of the three types, because they must recognise distinct molecular targets. There are more than 600 putative E3 ligases in the human genome (Li et al., 2008a). There are two main families of E3 ligase, the HECT family and the RING E3 ligases. The HECT family of E3 ligases transfer Ub to a target protein. The RING E3 ligases probably serve as recognition helpers for E2 to transfer Ub directly to the target (Weissman, 2001).

If the classical function of Ub is to tag proteins for degradation, it means the function of deubiquitinating enzymes (DUBs) is likely to be protein stability. Thus, based on the on the type of Ub chain that is attached to the substrate, deubiquitinating enzymes can start to react with target proteins. There are three general roles for DUBs. First, DUBs are important for Ub maturation once synthesised, since they can be synthesized in poly Ub chains or attached to smaller peptides. Secondly, DUBs can reverse the ubiquitylation action that targets proteins for degradation and in the process maintain Ub homeostasis through Ub recycling. Thirdly, DUBs can alter the type of Ub modification of a target enzyme (Komander et al., 2009). The human genome contains about 90 DUBs, but with only about 80 considered to be active (Clague et al., 2012). These proteins have been classed into 5 families depending on the motif architecture, UCH, OTU, MJD JAMM, and USPs. The USPs comprise the largest family DUBs with 63 members, and this family is considered to be responsible for cleaving Ub to prevent protein degradation by the proteasome. USPs have an important role to counteract the Ub ligases. In *S. pombe* there are 20 DUBs, of which Ubp5 is of interest to us due to its interaction with Hid3 (Kouranti et al., 2010a).

1.2.2 Ubiquitylation processes in the endomembrane system

1.2.2.1 ER-associate degradation of proteins

It is interesting that most of the proteins that are ubiquitinated in a cell are newly synthesized proteins just coming off of ribosomes (Kim et al., 2011). It is thought that ubiquitylation may serve as a quality control check system for newly synthesized proteins (Chhangani et al., 2012). Within the ER there is also a quality control system that makes sure that only properly-folded, functionally-active membrane proteins are transported onto the Golgi.

It is thought that glycosylation of proteins and insertion in to the membrane makes them much more resistant to deactivation and degradation so a screening system is required to remove unfunctional proteins. This is called the ERAD system (Kornitzer and Ciechanover, 2000) and associated with the Unfolded Protein Response. In humans, unproperly folded proteins appear to be recognized by molecular chaperones, such as the heat shock proteins (HSPs) and marked by phosphorylation by three particular membrane associated kinases PERK, IRE1, and ATF6. This slows down protein translation at the ER and permits targeting of unfolded proteins by an unknown E3 Ub ligase (Buchberger et al., 2010). Ubiquitylation is the signal that the protein is to be removed from the ER membrane and sent to the proteasome for degradation. In order to help decide between which proteins are to sent for degradation and those that are kept is discrimination made by the E3 ligase, but fine-tuned by the ER resident DUBs (Zhang et al., 2013b). Except for certain components, for example the chaperones and E3 ligases used, the ERAD system is conserved in the yeasts and the study of these systems has provided much insight into the animal systems (Buchberger et al., 2010).

1.2.2.2. The SREBP signalling processes of the ER and Golgi.

The control of cholesterol levels in human cells is mediated by the Sterol Receptor Element Binding Protein Pathway (SREBP). When cholesterol levels are low, SREBP is released from the endomembrane system to activate gene involved in cholesterol biosynthesis, of which there are about 30 in humans (Espenshade and Hughes, 2007). In humans, there are two SREBP genes that encode three proteins. Each protein controls the expression of genes to make a particular class of fatty acid or cholesterol. Under conditions of high cholesterol, SREPBs are held in the ER, but under conditions of low fatty acid/cholesterol, the SREBPs are released to migrate to the Golgi. In the Golgi, the two Site proteases Site-1 and Site-2 cleave SREBP releasing the N-terminal portion to enter into the nucleus and activate the transcription of target genes (Todd et al., 2006). The SREBP system is conserved in *S. pombe*, where it regulates the synthesis of the sterol ergosterol (Hughes et al., 2005). Interestingly, it is also the main system for low oxygen sensing and the regulation of hypoxia induced gene expression (Todd et al., 2005). In both humans and *S. pombe*, the cleavage of the N-terminus of SREBP (Sre1 in *S. pombe*) is initiated by their ubiquitylation by Golgi-resident E3 ubiquitin ligases (Stewart et al., 2011).

1.3 Human diseases associated with Golgi malfunction

The basic function of the Golgi complex is to locate and transport of proteins within and out of the cell. It receives proteins and lipids through the *cis*-face, they are modified as they traverse the Golgi and they exit through the *trans*-face. The failure for this process to work correctly has been associated with a number of human diseases. Alzheimer's disease is a brain disorder due to the accumulation of Beta-amyloid plaques in brain cells. It has been observed that disruption of Golgi structure comes before the accumulation of protein plaques (Joshi et al., 2014). They report that Aβ signals Golgi fragmentation, which in turn increases APP trafficking leading to accelerated Aβ production in a feed-forward loop. The amyloid-beta (Aβ) also raises the levels of cell cycle kinase, Cdk5, which phosphorylates GM130 leading to Golgi dispersal (Sun et al., 2008). Therefore, they concluded that there are two ways to reverse or prevent the harmful effects of beta-amyloid secretion in AD patients, first by inhibiting or a mutated *Cdk5* or degrading increasing APP protein degradation through the ubiquitin-proteosome pathway. The disruption of Golgi structure has also been implicated in other neurological disorders, such as Parkinson's and Creutzfeldt-Jakob diseases (Gonatas et al., 2006).

The previous example demonstrated how signalling processes can disrupt Golgi function. There are many examples where altered function of a Golgi protein itself results in a severe disease (Bexiga and Simpson, 2013). This would be expected for a process deciding the localisation of so many numbers of proteins. In addition to that, there are many ways by which the Golgi can be affected. For example, a defect in retrograde protein transport may have as severe effects as a disruption to anterograde transport. Over the past few years, many diseases have been linked to the malfunction of the Golgi complex, including skin diseases, neurological disorders, skeletal dysplasia and connective tissue disorders. I will briefly mention some of them here. An example of altered protein trafficking for a skin disorder is Dyschromatosis universalis hereditaria, which commonly occurs in Japan (Zhang et al., 2013a). A defect in the gene *ABCB6* gene leads to accumulation of its translated protein in the Golgi causing malfunction. An example of neurological disease comes from a defect in the myelin gene, *PLP1*. A mutation in *PLP1* causes a rare disease known as Pelizaeus-Merzbacher syndrome. This disease results from an accumulation of the misfolded PLP1 protein remaining in the ER resulting in a defect in retrograde transport of proteins from the Golgi to the ER, contributing to Golgi fragmentation (Inoue, 2005). Another Neurological disease *proximal spinal muscular atrophy* is caused by low expression of the *SMN1* gene leading to accumulation of SMN granules in the late stage of Golgi network and stopping vesicle transport (Ting et al., 2012). Also, *North Sea myoclonus* neurological disease is where a mutant cannot localise the protein to the *cis*-Golgi. This has

been linked to the *X-linked mental retardation 1 (MRX1)* and is caused by mutation in the *IQSEC2* gene located on chromosome Xp11.2 (Boissé Lomax et al., 2013; Corbett et al., 2011). Many malfunctioning proteins found to cause diseases have been classified under connective tissue disorders. In these cases, disrupted traffic through the Golgi does not all proper delivery of connective tissue proteins to the extracellular matrix. Examples of these are *RIN2* syndrome, *MACS* syndrome, Gerodermia osteodysplastica, Menkes disease and Cutis laxa (Bexiga and Simpson, 2013). One final example for a possible link of skeletal diseases to Golgi function is through the uncharacterized transmembrane protein *TMEM165* (Foulquier et al., 2012). The wild-type protein is in the Golgi complex and to the endosomal-lysosomal system. In human cells lacking this protein, trafficked protein are spread to different locations around the cell. It appears that *TMEM165* has important role in Golgi glycosylation and Golgi morphological structure.

1.3.1 Skeletal dysplasia

There are two causes of the skeletal dysplasia Dyggve-Melchior-Clausen syndrome (DMC) that I will discuss. The first is caused by mutations that alter or eliminate the function of the RAB33B and the second is the protein DYMECLIN (DYM). DMC syndrome is characterized by short stature due to skeletal dysplasia deformities accompanied by microcephaly and intellectual delay (Dyggve et al., 1962). DMC is very similar to another genetic syndrome called Smith-McCort dysplasia (SMC). Smith-McCort dysplasia is distinguished from DMC by the lack of mental retardation, but the skeletal deformities are similar (Khalifa et al., 2011). Both DMC and SMC are autosomal recessive disorders in that in the disease occurs when a child inherits
mutant alleles from each parent and so is homozygous for mutations at that locus. Heterozygous individuals do not appear to be affected by the disease.
parent and so is homozygous for mutations at that locus. Heterozygous individuals do not appear to be affected by the disease.

This syndrome has been reported in many countries around the world and early diagnosis is important. Preventing the disease by counselling parents to avoid passing the disease on to children is the only current means of preventing the disease, since it is a standard recessive genetic disease (Bayrak et al., 2005; Khalifa et al., 2011). Early treatment is important to alleviate some of the aggravation of the disease, such as using surgery to join vertebrae called spinal fusion or other surgical techniques to correct obvious skeletal abnormalities. Nowadays, pre-implantation genetic diagnosis can be used for those parents who appear to be at risk to pass the syndrome to their children. Although it is rare syndrome, it appears that the

number of cases continues to increase. In 2002 there were only 58 cases known and this increased to 90 cases in 2004 (Kandziora et al., 2002) with 15 cases attributed to nine unrelated families in Egypt with both males and females being affected by a defective *DYM* gene (Santos et al., 2009).

1.3.1.1 Malfunctioning RAB33B as a cause of SmithMcCort dysplasia

RAB33B was the second locus found where mutation led to the DMC syndrome (Alshammari et al., 2012). RAB33B resides on chromosome 4 at position 4q31.1 and homozygosity of mutation leading to complete malfunction of the protein is the cause of DMC. There are around 70 different RAB proteins in the human genome and they have been found in a variety of organelles (Bem et al., 2011; Liegel et al., 2013). RAB proteins are small intracellular GTPase involved in cellular protein trafficking. Seven of them are involved in Golgi structural maintenance, Rab1, Rab2, Rab6, Rab18, Rab33B, and Rab43. These RAB proteins likely serve to recruit other factors to Golgi membrane. Like other GTPases, such as the Ras and Rho GTPases, these cycle between a GTP-bound and a GDP-bound form. In the GTP bound form they are attached to the Golgi membrane to carry out their recruitment function. Rab GTPases have low GTPase activity, but are activated by guanine nucleotide exchange factors (GEFs) (Martínez-Alonso et al., 2013), which is likely their mechanism of regulation. The Rab proteins are believed to be the link between Golgi function and the regulation of protein transport (Liu and Storrie, 2012) and the identification of loss of RAB33B leading to DMC is strongly supports this. RAB33B is specifically thought to be involved in retrograde protein transport through the Golgi, which also shows that anterograde and retrograde movement are closely linked (Dupuis et al., 2015). Several other human diseases are caused by *RAB* mutations including cancer. For example *RAB18, RAB1GAP* and *RAB2GAP* have been implicated Warburg Micro syndrome (Bem et al., 2011; Liegel et al., 2013) and *VPS53* in progressive cerebello-cerebral atrophy type 2 (Feinstein et al., 2014).

1.3.1.2 Lack of DYMECLIN function resulting in DMC syndrome.

The *Dymeclin (DYM)* gene responsible for DMC maps to chromosome position 18q12-21.1 (El Ghouzzi et al., 2003). The name DYMECLIN comes from the combination of the researchers from Denmark that first systematically reported the disease Dyggve, Melchior and Clausen (Dyggve et al., 1962). This gene encodes a peripheral membrane protein located in the Golgi (Dimitrov et al., 2009) that is involved in protein trafficking (Osipovich et al., 2008). The DYM protein appears to be able to shuttle back and forth between the cytosol and Golgi, thus it

is likely tethered to the Golgi by fatty acid modification of the protein (Denais et al., 2011; Dimitrov et al., 2009). The cases of DMC that have been shown to involve DYMECLIN result from mutations in the gene that eliminate the protein (Khalifa et al., 2011). In cells of DYM patients there were ultrastructural analomolies to organelles within cells, but the Golgi apparatus appeared normal (El Ghouzzi et al., 2003).

1.4 Cancer as a genetic disease

The human body contains a huge number of cells and every cell, with the notable exceptions of red blood cells and platelets, contain a nucleus, where the majority of them contain 23 pairs of chromosomes, one of each inherited from the mother and father. As mentioned above for the cases of heritable genetic diseases, a pre-disposition to cancer can also be passed from parents to children. Every cell in our body has about 20,000 to 25,000 genes, of which there have been 20203 annotated proteins (www.expasy.org). This is likely to be a large underestimate of the true number of functional genes if non-coding RNAs are considered (Griffiths-Jones, 2007). In any cell type, around half of these are activate to make a certain cell type and of this half, two different cell types differ in protein-coding genes expressed by about 2000. Thus each cell probably is different by about 2000 proteins.

Cancer is a disease when the gene expression program in cells gets out of whack. Cancer can start from the gene expression mis-programming of one cell and the replication of this cell could lead to abnormal growth of cells in a group called a tumour. Once started, it can spread to other parts, called metastasis. Thus, the healthy tissues will be affected by these cells with over time either by the tumour growing and changing the local tissue environment or spreading to new tissues (Bertram, 2001). There are about 15 million new cases of cancer each year (Ferlay et al., 2014). Approximately, there are 200 types of cancers, and based on that it is classified as heterogeneous disease which may causes by different kind of factors. The most common types of cancers are breast, lung, colorectal and prostate (Ferlay et al., 2014). Many studies indicate that cancer caused by both environmental and genetic (heredity) factors. Many genetic mutations do not affect cells and cells have the ability to repair the majority of them. Not all mutations have the same effective as some prevent proteins from being made, and others alter the function of the protein made, or a mutation causes more protein made than required. Genetic mutations can occur by different ways, such as point mutations, deletions, and insertions. Point mutations are the most common, where one base is changed to another. The effect of a mutation is also dependent on the type of cell in which the mutation occurs, for example, reproductive cells going through meiosis or somatic cells undergoing mitosis (Wilkins

and Holliday, 2009). Mutations occurring in germ cells called gametes (sperm or eggs) are called hereditary or germline mutations, which is the cause of about 5 to 10 % of all cancers.

However, specific mutations within germline cells are not necessarily sufficient to initiate cancer formation. Cancer does not appear at once, and more than one mutation is needed to start tumorigenesis (Kushi et al., 2006). Nevertheless, there are certain genes in which inherited mutations will cause a predisposition to cancer, if an opportunistic mutation occurs later. These types of genes are known as proto-oncogenes, such as RAS that when activated uncontrollably causes increases in hormonal signalling responses, the tumour suppressor genes p53 and pTEN, genes for DNA repair and many others (Lee and Muller, 2010). Mutations in the protein p53 have been implicated in more than 50% of all cancers (Hollstein et al., 1994). In most cases, a mutation in one of these genes is not critical, because the allele on the other chromosome makes a normally functional protein, called heterozygosity. As cells get older, it is possible to get a mutation in the healthy allele that disrupts a critical process, such as in one of those above, that leads to cancer formation. This is called Loss of Heterozygosity (LOH), where the second allele was already inactivated. These mutations can come from a number of external sources generally known as environmental factors. Protein like p53 and pTEN function as tumour suppressor proteins and are encoded by what are called Tumour Suppressor Genes (TSGs). The have various functions from activating repair pathways following DNA damage and initiating apoptosis, like p53, to acting as inhibitors, like pTEN, of signalling pathways, such as RAS-mediated signalling pathways, that might get out of hand. TSGs can be involved in stopping the cell cycle upon DNA damage in order to allow repair to get started. Loss of TSG function can lead to reduced apoptosis and then increased tumour growth, but both alleles of the gene in the cell need to be mutated to inactivate TSGs.

Environmental factors are more likely to cause cancer than genetic factors. Unlike genetic factors, we can avoid these environmental factors, such as by doing exercise, reducing alcohol consumption and maintaining a healthy diet (Bertram, 2000). Other factors, like hormones, UV/isotopic radiation, viruses, bacteria and chemical substances are commonly encountered as they are found in air, water and food. An example of how environment can determine the incidences of cancer in a population is the migration of Asian people from the far-east to America. The far east diet and little industrial/agricultural exposure led people to be classified as low-risk of prostate and breast cancer, but incidences have increased enormously in Asian populations since their immigration to America (Kushi et al., 2006). A major cause of cancer is the damage to DNA resulting from exposure to an environmental factor. The damage to DNA is

a major cause of the mutations discussed above. Therefore, cells have developed very complex mechanisms in order to repair DNA. For example, if a double-stranded DNA break occurs there are a number of signalling pathways that are started to induce the DNA damage response. In humans, the *ataxia telangiectasia mutated* (ATM) kinase is induced to phosphorylate the checkpoint kinases CHK1 and CHK2 and p53. This stop in the cell cycle allows time for the DNA repair process to be initiated. If the checkpoint is stopped for too long, then apoptosis is initiated (Shiloh, 2003). The DNA-PK pathway is also induced to initiate non-homologous end-joining repair pathway, but it can also phosphorylate p53 to help induce its activity. It is interesting that activation of the DNA-PK pathway appears to have big effects on Golgi function. DNA-PK directly phosphorylates the Golgi protein GOL3PH leading to Golgi breakup (Farber-Katz et al., 2014). This GOL3PH-mediated Golgi dispersal appears to be essential for cells to survive DNA damage. However, over-expression of active, phosphorylated GOL3PH can prevent cellular apoptosis by DNA damage, therefore, it provides a good example how cell activation of a protein can be oncogenic.

1.4.1 Ubiquitin and cancer

Ubiquitylation of proteins is one of the earliest responses of the cell to DNA damage (Brinkmann et al., 2015). PAR polymerase 1 at double-stranded DNA breaks attracts the E3 ligase CHFR that the ubiquitylates the protein leading to its dissociation and allowing DNA repair to begin. It has also been reported that ubiquitylation of histone H2AX is important, because it then attracts the ATM kinase to initiate the check in the cell cycle. As shown above, ATM kinase phosphorylates p53 in order to help activate it. The activity of p53 is kept in check by ubiquitylation of p53 by the E3 ligase MDM2, which targets p53 for degradation by the proteasome (Momand et al., 1992). Another E3 ligase COP1 destabilises p53 by ubiquitylation of p53 and targeting for degradation (Dornan et al., 2004). A major part of p53 stabilisation by ATM is the phosphorylation of both MDM2 and COP1, which leads to their inactivation. There are many other E3 ligases that appear to regulate p53, such as Cul4B, E6-AP and Cul4A-DDB1, as well as E4 ligases that catalyse the polyubiquitylation of proteins (Brinkmann et al., 2015). It follows that p53 is activated by deubiquitylation and this is discussed in more detail below regarding possible functions of HID1. pTEN is also a TSG that is regulated by the ubiquitylation/deubiquitylation cycle and this is also discussed below. Therefore, it is not surprising that mutations in genes involved ubiquitin-related process had been implicated in the initiation of cancer.

1.5 Environmental stress signalling

Stresses can be defined as external factors or conditions that affect a biological system. There are various types of environmental stresses, but are categorised into two classes, biotic and abiotic stresses. Multicellular organisms, such as animals and plants face both. Common biotic stresses would be caused by bacterial or viral infection, with insect or carnivore wounding posing an added stress to plants. Microorganisms are mainly bothered by biotic stress as abiotic stress on any one cell, perhaps with exception of non-virulent virus infection, a biotic stress is likely to cause the death of the cell. Abiotic stresses include oxidative stress caused reactive oxygen species in an oxidising environment or lack of oxygen in a hypoxic environment, chemical stress through heavy metals or pollutants, temperature stress, radiation stress, or osmotic stress. The key to an organism being able to survive a stress is for it to detect the stress and to respond to it in time. Cells have gained very complex signalling mechanisms in order to respond to environmental stresses. Two good examples of stress signalling pathways are the ATM and DNA-PK mediated pathways caused by chemical or radiation induced damage to DNA. Cells of both multicellular and microorganisms need cell signalling mechanisms that start outside the cell at the surface of the plasma membrane. The mitogen-activated protein kinase (MAPK) signalling pathways are a family of signalling pathways that allow for sensing and signalling outside stresses. There are four members of the family, the extracellular-signal regulated kinase (ERK) involving p42/44 MAP kinase, the stress activated protein kinase pathway 1 (SAPK1) involving JNK, the SAPK2 pathway involving p38 MAPK kinase, and the Big MAP kinase pathway involving ERK5. The ERK pathways respond to growth signals and mitogens and the SAPK pathways respond to external stresses (Koul et al., 2013). Parts of the MAPK signalling pathways conserved among all eukaryotic cells and regulate a variety of cellular functions, including gene expression, cellular homeostasis, and differentiation in response to different environmental stimuli (Herskowitz, 1995). MAP kinases are usually activated by specific MAP kinase kinases and those by MAP kinase kinase kinases that respond more specifically to signals. The induction of MAPK signalling pathways by specific upstream kinases appears to be the mechanism that separates stress signalling responses. However, there is cross-over in signalling mechanisms and this is much more in single cell eukaryotes like yeast. The end result of MAPK signalling is modification of transcription factors that start gene expression programs that allow the cells to respond to the stress. Stress signals can come from within cells from organelles that are disrupted in function (Kim and Choi, 2010), such as the initiation of MAPK signalling pathways in response to ER stress (Darling and Cook, 2014; Kim and Choi, 2010).

In *S. pombe*, the homologous pathway to the human ERK1 and ERK2 pathways is the Pmk1 MAPK pathway that is also known as the cellular integrity MAPK pathway (Garcia et al., 2009). This appears to be an important pathway for responding to extracellular stress signals. There is also the Spm1 MAPK pathway that appears to link environmental signals to cell morphology (Zaitsevskaya-Carter and Cooper, 1997). An important MAPK signalling pathway that is induced by a number of stresses is known as the Sty1-mediated MAPK pathway. Sty1 is the homolog of the human p38 MAPK and the Hog1 MAPK of *S. cerevisiae*. It is phosphorylated by the MAPKK Wis1 where it translocates to the nucleus to activate the transcription factor Atf1 to induce transcriptional responses (Nguyen et al., 2002).

With *S. pombe* having similar stress signalling processes to humans there has been a lot of interest in knowing how gene and protein expression changes in response to stress. Using microarrays, the transcriptional responses of the fission yeast *Schizosaccharomyce pombe* were analysed in cells exposed to five different stresses, hydrogen peroxide, heat, the heavy metal cadmium, the osmotic stress compound sorbitol, and the DNA damaging chemical methylmethane sulfonate (Chen et al., 2003). They identified two different types of genes changing in response to stress, a core set of genes that they called the core environmental stress response (CESR) genes and sets of genes changing for each specific stress, call the stress-specific environmental stress response (SESR) genes. The CESR genes are mainly dependent on Sty1 and Atf1 and the SESR genes have specific induction mechanism. From these early studies, global ‘omics’ studies have expanded to look at post-transcriptional control (Mata et al., 2005), changes in both the transcriptome and proteome (Schmidt et al., 2007) and in the metabolome (Pluskal et al., 2010). A study to look at gene expression and protein changes in the transition from proliferation to quiescence in *S. pombe* showed that gene expression goes way down and the proteome changes too. However, instead of across-the-board changes, like for the mRNAs, protein levels are selectively changed and are controlled by cell size (Marguerat et al., 2012). In this study, they also determined that most mRNAs are present in 1 to 10 copies per cell, but proteins are present from hundreds to thousands of copies.

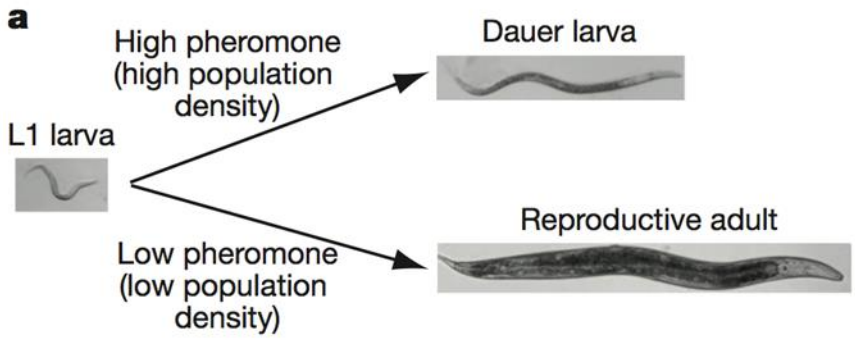


Figure 1.3. The alternative growth stages of *C. elegans*. *C. elegans* starts in a small larval stage. Under good growth conditions it grows into an adult stage and can reproduce (Low population density). Under conditions of high population density where food might become limiting a pheromone is emitted which causes it to form a quiescent growth form called a Dauer larva. The image was reproduced from (McGrath et al., 2011).

1.6. The HID1 protein

1.6.1. Human HID1

A potential paralogue of DYM in humans is HID1. The two proteins have been grouped into the same clan Golgi_traff (pfam CL0456) based on sequence and functional homology. The *HID1* gene lies on chromosome 17 in a region that has been well characterised as a site where loss of heterozygosity contributes to the development of cancer (Risk et al., 1999). The well-studied and highly publicised tumour suppressor gene BRCA1 lies in region 17q21 (King, 2014). HID1 lies nearby in the *TOC* region 17q25 that appeared to be associated with the onset of cancer also (Fukino et al., 1999). The gene that was thought to be responsible was eventually cloned and called *DMC1* for Down-regulated in Multiple Cancers (Harada et al., 2001). They classified *DMC1* as a Type II Tumour Suppressor Gene, because it appeared to be reduced in expression or the transcript missing entirely in a large proportion of cancers that they studied.

The gene is on Chromosome 17 between positions 72,946,839 to 72,968,900 bp making the gene just over 22 kb long. It has 19 exons and 18 introns. It encodes a protein of 788 amino acids (Maglott et al., 2011). General screening experiments have given an idea as to where the protein is located (Pilot-Storck et al., 2010; Keshava Prasad et al., 2009). It has been found in many different tissues and organs, heart, in skeletal muscle, colon, spleen, kidney, liver and lung. It was at high levels in the brain and placenta suggesting a particular importance of this protein in these tissues. DMC1 has two synonyms by which they are both commonly referred to in databases, c17orf28 and HID1. The HID1 term is more commonly used now due to its homology with the HID-1 protein of *C. elegans*.

1.6.2 *C.elegans* HID-1

C. elegans has become an important model system in molecular research due to its simplicity in handling and it has similarities to human physiology. About 35% of human genes have homologues in *C. elegans*, and *C. elegans* has similar physiological behaviours, such as development of Alzheimers disease, diabetes and even depression (Kaletta and Hengartner, 2006). *C. elegans* has a short life cycle, going through an embryonic stage, four larval stages (L1-L4) and adulthood. This multicellular organism goes through its 4 larval stages under good environmental conditions, such as suitable temperature, if there is enough food and in the absence of a pheromone indicating population density. Since scientists discovered that calorie restriction and mutation affecting feeding behaviour could lead to an increased lifespan, it has become an increasingly important model organism. Furthermore, *C. elegans* can adopt an

alternative growth form called the Dauer larval stage (McGrath et al., 2011), which is a reduced growth stage when environmental conditions become unsuitable or population density becomes to great (Figure 1.3). Normal growth can restart if population is diluted or food returned in the case of starvation-related Dauer formation (Fielenbach and Antebi, 2008).

The *hid-1* gene of *C. elegans* was found in a screen looking for mutants of the nematode which adopted the Dauer formation under good growth conditions, but with the temperature slightly raised (Ailion and Thomas, 2003). *hid-1*, which stands for high-termperature induced Dauer formation, is one gene of many that have been discovered controlling dauer larva formation. These genes can be classified into two types: Dauer formation defective (Daf-d) where larva cannot form Dauers under inducing conditions, and Dauer formation consistitutive (Daf-c) where Dauers forms spontaneously under condition where it would grow normally. The gene *hid-1* is classed as Daf-c. The Daf-c genes have become interesting due to their contribution to a pheromone sensing and signalling pathway analogous to Transformation Growth Factor Beta in humans, which plays a vital role in growth of breast cancers (Patterson and Padgett, 2000). The protein DAF-7 is a ligand for the receptors DAF-1 and DAF-4 to activate the transcription factors DAF-8 and DAF-14 to up-regulate genes for Dauer formation. The DAF-D proteins appear to form a system that is the antagonist of this pathway (Patterson and Padgett, 2000). Many other signal transduction pathways play vital role regulating dauer formation, because this stage can form under other poor growth conditions, such as high temperature or nutrient limitation. A good example is the gene *daf-11*, which is a (Daf-c) gene that encodes an quanylyl cyclase that appears to activate cGMP-gated channels involved in neuron development (Coburn et al., 1998). cGMP-gated channels have a role in thermo sensation and dauer larval development (Birnby et al., 2000). A particular relevant example to my study is the function of the DAF-16 protein in insulin and oxidative stress signalling (Dillin et al., 2002). DAF-16 is the orthologue of the human FOXO4 transcription factor involved in insulin signalling (Nakae et al., 2008). In *C. elegans*, insulin binds to its receptor activating the AKT kinase which phosphorylates DAF-16, thereby activating it. In humans, it is thought that deregulation of the AKT pathway via FOXO4 activation may play a key part in the development of diabetes. Interestingly, the *daf-16* mutation, which is a Daf-d type, directly suppresses the *hid-1* temperature induced Dauer formation phenotype (Ailion and Thomas, 2003). This suggested some type of antagonistic interaction between Hid-1 and DAF-16, and this could translate into a similar interaction between HID-1 and FOXO4 in humans. A potential mechanism for this has been developed through studies of both *C. elegans* and *S. pombe*.

Most of what is known about HID1 function comes from studies on *C. elegans*. HID-1 has been shown to be a peripheral membrane protein associated with the *medial-* and *trans-*Golgi apparatus (Wang, L. 2011), but it has also been found in dense core vesicles (DCV) and at the plasma membrane in the synaptic region of neurons (Mesa et al., 2011). HID1 is highly expressed in neurons compared to other tissues in the nematode. However, like DYM, HID1 was found to dynamically associate with the Golgi membrane and its attachment is mediated by myristoylation (Mesa et al., 2011). Two groups independently came to the conclusion that HID1 acts to sort neuropeptides in neuronal development and communication (Mesa et al., 2011; Yu et al., 2011). The study of Yu et al., (2011) using *C. elegans* investigated more closely *HID-1* function in the DCVs. DCVs are formed in the *trans*-Golgi network (*TGN*), where different types of vesicles are generated. Many steps have to be carried out to convert immature *DCVs* to mature *DCVs* to be localised at the plasma membrane in order to release their proteins, peptide neurotransmitters, growth factors, amines and hormones. However, in the *hid-1* mutant the *DVCs* produced never appeared to develop to maturity. While the luminal peptide cargoes of DCVs were reduced, the DCV membrane cargoes were increased. The researchers found that the reduction in peptide cargoes occurred because of lysosomal degradation. Therefore, it appears that HID-1 play a vital role in regulating DCV traffic by preventing the mis-sorting of peptide cargo from the plasma membrane to the lysosome.

1.6.3. HID1 in *S. pombe*

The functional orthologue of HID1 in *S. pombe* appears to be the protein Ftp105. I say functional orthologue because *S. pombe* actually has three HID1 orthologues. SPAP27G11.12 and SPBP19A11.07c are listed in PomBase as human HID1 orthologs, but they have not as of yet been published. However, a global study of protein localisation has reported that they might be located in the cytoplasm or the nucleus (Matsuyama et al., 2006). The relationship between the three paralogs I describe in more detail in Chapter 3. Ftp105 hasn't been the subject of a particular study, but some of its function has come to light from studies investigating the function of the DUB family members in *S. pombe*. *S. pombe* is a good model for studying certain protein families like the DUBs because they have a limited number compared to humans, 20 compared to 96 (Kouranti et al., 2010a). Through tagging with GFP, it was observed that the Dub Ubp5 localises to the Golgi membrane *S. pombe* and this is due to interaction with Ftp105. Ftp105 was also shown to localise to the Golgi through GFP tagging.

In the Ftp105 mutant, Ubp5-GFP fluorescence becomes diffuse due to redirection to the cytosol. In addition, through an interaction study using immunoprecipated Ftp105 through TAP-tagging, a physical interaction between Ftp105 and Ubp5 was confirmed (Kouranti et al., 2010). It was speculated that the function of Ftp105 may be to sequester Ubp5 to the Golgi to regulate its function. A subsequent study investigating the phosphorylation status of the *S. pombe* Dubs indicated that phosphorylation of Ubp5 was necessary to promote the association with Ftp105. By this association, Ftp105 may play a role in ubiquitination processes in *S. pombe* and by analogy HID1 in *C. elegans* and humans.

Ubp5 is interesting, because it is the *S. pombe* orthologue of the human protein USP7/HAUSP. FOXO4 transcriptional activity is rapidly increased in response to oxidative stress by ubiquitylation (van der Horst et al., 2006). Its ubiquitylation is the signal for the association of FOXO4 with USP7 in order to decrease transcriptional activity. Interestingly, lower expression of FOXO4 appears to promote gastric cancer cell line growth and increase tumour metastasis (Su et al., 2014). USP7 also plays an important role in cancer development or its prevention through two well-known tumour suppressors, p53 and pTEN. PTEN stands for Phosphatase and tensin homolog and mutation in the gene inactivating the protein have been implicated in the development of cancer. In cells that have lost pTEN phosphatase activity, the phosphoinositol-3-Kase (PI3K) signalling pathway is up-regulated leading to cancer progression. This is one of the most common ways that cancer forms (Shaw and Cantley, 2006). In addition, the nuclear location of pTEN is critical to the inhibition of PI3K pathway activity, which is controlled by ubiquitin. Ubiquitylation of pTEN causes it to move from the nucleus to the cytosol and so inhibiting its function as a tumour suppressor (Song et al., 2008). Therefore, more USP7 in the cytosol/nucleus would maintain PI3K inactivation and tumour progression. The protein p53 is a strong tumour suppressor through activation of apoptotic signalling pathways. Ubiquitylation by USP7 can influence p53 in a few ways, including deubiquitination of the complex that keeps it associated with DNA to for transcriptional activation (Zhao et al., 2004), and it can directly deubiquitylate p53 in order to keep it active (Li et al., 2002). Therefore, in contrast to the pTEN system, deubiquitylation would lead to more tumour suppressor action.

There are other examples of potential Golgi function could lead to cancer when not properly functioning, such as through the Ras GTPases. Ras is linked to mTOR signalling, which involves pTEN, through the mitogen-activated protein kinase pathway. Mutations that activate Ras or the sustained excitation of Ras signalling pathways can promote tumour growth

(Shaw and Cantley, 2006). Not only are there Ras GTPase present in the Golgi, but Ras localisation within the cell governs their function, which is dependent in many cases on Golgi-mediated sorting (Chang and Philips, 2006; Onken et al., 2006). Furthemore, Ras activity can be affected by ubiquitinylation (Jura et al., 2006).

1.6.4 The HID1 ortholog in *S. cerevisiae*, ECM30

By sequence, the most similar protein to HID1 in *S. cerevisiae* is called ECM30. It is still referred to in the Saccharomyces Genome database (yeastgenome.org) as a putative protein of unknown function. The gene is located on Chromosome XII between nucleotide positions 1007421 and 1011245 and the CDS is 3825 nucleotides long. It encodes a protein of 1274 amino acids, which is larger than Ftp105 at 872 nucleotides long. ECM30 mutants show reduced fitness compared to wild-type strains (Giaever et al., 2002) and it has abnormal bud morphology (Watanabe et al., 2009). It has abnormal vacuolar morphology in the presence of 0.4M NaCl (Michaillat and Mayer, 2013). It is sensitive to a number of different stress treatments, such as hydroxyurea and heavy metals. These phenotypes can certainly be associated with disrupted protein sorting by the Golgi apparatus. A study by (Costanzo et al., 2011) to map the genetic interactions for the budding yeast, *S. cerevisiae* found that the proteins *ECM30, UBP15*, and *PAR32* have a similar interaction to those genes involved in the *GAP1* sorting pathway. The gene *GAP1* encodes an amino acid permease present in the plasma membrane and so it is sorted by the Golgi trafficking pathway. Its levels are regulated by degradation in the vacuole, and like many plasma membrane proteins, this endocytosis and vacuolar degradation is started by ubiquitination of the protein (Soetens et al., 2001). They validated the interactions for ECM30 by deleting each gene, *ecm30*Δ*, ubp15*Δ*,* and *par32*Δ mutants which cause mislocalization and less activity of the *Gap1* permease phenotypes (Costanzo et al., 2010).

1.7. *S.pombe* as a model system

The yeasts are eukaryotic organisms that belong to the kingdom of fungi. This kingdom ranges between 1.5 to 5 million species comprising filamentous fungi, mushrooms, lichens, unicellular yeasts and both animal and plant pathogens. *S.pombe* is a member of the Eucarya classified under a phylum of *Ascomycota* that contains well-known species, such as *Saccharomyces cerevisiae* (baker's yeast), Aspergillus nidulans, and *Neurospora crassa*. It is the largest phylum within the fungal kingdom with around 65,000 species. Fungi share recent ancestor with animals making them more related to animals than to plants. A phylogenetic clade

named *Opisthokonta* refers to the flagella that are believed to be the common ancestor of animals, fungi and protists (Adl et al., 2005; Koonin, 2010). The taxonomy of *S.pombe* according to NCBI is the following lineage: Eukaryota/ Fungi/ Ascomycota/ Archiascomycetes/ Schizosaccharomycetales/ Schizosaccharomycetaceae/ Schizosaccharomyces. It evolved from *S. cerevisiae* about 1000 million years ago.

S.pombe is a unicellular fungal model organism containing three chromosomes with a small genome consisting of 13.8 million base pairs and about 4,824 genes. In terms of gene numbers it is the smallest eukaryote known (Wood et al., 2002). It is a cylindrical cell with a diameter of 3-4 μm and a length between 6 and 14 μm. Starvation can cause the cells to shorten making them look a lot like *S. cerevisiae*. The normal cell state of pombe is haploid, but it can be a stable diploid, and it has a doubling time of between 2 and 4 hours. Under conditions of stress zygotes are formed. *S. pombe* has many advantages as a model organism, such as fast growth on a variety of well-defined media (Yanagida, 2002). It has very few duplicated genes, making it a good model for functional genetics. Laboratory strains are believed to be predominantly isogenic, avoiding many of the problems encountered with studies in *S. cerevisiae*. But like *S. cerevisiae*, it is easily transformable in order to make mutants or heterologously express genes with many genetic tools, such as selective marker genes and transformation cassettes. Many groups of genes are found conserved in *S. pombe* and humans, but missing in other model organisms. Most importantly, it shares many features with higher eukaryotic cells, such as molecular, genetic and biochemical similarities, which can help us to study homologous genes functions and processes conserved in other organisms, such as cell cycle control. Another important advantage of studying *S.pombe* is that its cellular signaling pathways are similar to those of humans (Wood et al., 2002). It has been used mainly to study cell cycle control, mitosis, meiosis, chromosome dynamics, epigenetics, DNA repair and mRNA processing. Nowadays, *S. pombe* is playing a big role in post-genomics studies, because mutants can be used for RNA sequencing and proteomics approaches.

1.7.1 The pombe cell cycle

The main aim of the cell cycle is to transfer genetic information from a parent to two daughter cells by going through four stages: M (mitosis), G_1 (first gap phase), S (DNA synthesis) and G_2 (second gap phase) in order to create two daughter cells (Sherr, 2004). The cell cycle progression is regulated by two important families of proteins the cyclin proteins and catalytic cyclin-dependent kinases (CDKs). In all eukaryotes, yeast to human, the cell division is

controlled by the cyclin/CDK complex. There are different CDKs in *humans*, while in *S. pombe yeast* there is only the one CDK Cdc2 that interacts with many cyclins depending on the stage of the cell cycle (Malumbres and Barbacid, 2009). The interaction of Cdc2 with other cyclins is a check to ensure the expression of genes necessary to carry the next phase (Novak et al., 1998). Cdc2 and interacting partners, such as cyclin Cdc13, are regulated through a negative-positive interaction through phosphorylation by Wee1 kinase, and dephosphorylation by Cdc25 phosphatase. A study of gene expression throughout the cell cycle showed that there are groups of gene expressed during each of the phases and that each phase appears to be regulated by an important TF or TF complex (Rustici et al., 2004). Cell cycle control is not only genetic and must respond to the environment to make sure conditions are correct for cell division. The cell cycle and cell division can be stopped for example by nutrient limitation. Stress can also regulate the cell cycle. The MAPK stress signalling pathway through Wis1 and Sty1 are important for initiating the entry into Mitosis (Parr, 2012; Shiozaki and Russell, 1995). However, both Wis1 and Cdc2 are necessary for cells to enter and keep a quiescent cell state (Sajiki et al., 2009). Finishing the cell cycle once started and stopping the cell cycle involve a complex interaction of genetic components (Novak et al., 1998) and *S. pombe* represents an important model to understand the regulation of the cell cycle and identifying how cells grow and differentiate, such as cancer cells (Malumbres and Barbacid, 2009).

1.8 Proteomics

Proteomics is the study of the proteome, which has been said to be the next step from transcriptomics and refers to the large-scale analysis of proteins (Pandey and Mann, 2000). The proteome can be defined as a complete set of proteins expressed by a cell (Wilkins et al., 1996). Proteomics initially started with protein extracts being separated on 2-dimensional gels in order to separate the mass of proteins within a cell (Rogowska-Wrzesinska et al., 2013). 2-D gels can be used to quantify relative amounts of proteins in cell samples through the use of protein labeling, such as DIGE, which labels proteins with fluorescent dyes, much like microarrays. Nowadays, gel-based proteomics have given way to mass-spectrometry (MS) based proteomics, particularly for being able to quantify proteins for sample comparison (Schulze and Usadel, 2010; Washburn, 2011). The main difference between gel proteomics and MS proteomics is that MS recreates proteins from peptides, where quantification of proteins on 2-D gels comes from the entire protein and identification comes from the peptides. MS has played an important role in the development of protein quantification. Because it is fast and

sensitive, it has become become the method of choice for protein identification and quantitation (Patterson and Aebersold, 2003).

There are both protein-labelled techniques and label-free techniques for quantifying proteins. Labeling techniques can be done by comparing proteins from cells grown on a heavy isotope of an element found in proteins to cells grown on the corresponding light element, such as $^{15}N/^{14}N$ (Kussmann et al., 2008). The peptides of trypsin-digested protein extracts can be tagged with a compound detectable in the mass spec and quantified that way, called isobaric tagging like the iTRAQ technology (Gygi, 2000). The technique becoming more and more common is a label-free technique that counts the number of times a peptide from a digested protein extract is counted in the mass spectrometer (Kussmann et al., 2008). The general workflow in a label-free experiment is to isolate the proteins from the tissue and this can be done in ways to get only the soluble proteins, using aqueous extraction techniques or all proteins using non-aqueous solvents. The proteins can either be purified on acrylamide gels and digested with trypsin or directly digested in a tube to give the peptide mixture. The peptide mixture is then injected onto the mass spec using some type of separation device, such as liquid chromatography. A number of software programs have been developed to measure masses, count the peptide hits on the MS and convert this data into protein amounts (Langella et al., 2013; Valot et al., 2011).

There have been some proteomic studies conducted on *S. pombe* and the number is likely to grow in the future. As stated above, a proteomic approach was used to estimate that proteins range in amount from hundreds to thousand copies in a cell (Carpy et al., 2014; Marguerat et al., 2012), where mRNA levels were 10-100 times less (Marguerat et al., 2012). A study of the proteome of the *sty1Δ* mutant in response to oxidative stress confirmed the complexity of Sty1-mediated signalling and demonstrated that there is little correspondence between proteins levels and gene expression in response to stress. In contrast, a comparative study of the *S. pombe* proteome and transcriptome showed quite a bit of correlation between transcript and protein levels, especially for signalling and metabolic proteins (Schmidt et al., 2007). What makes proteomics a particularly interesting tool is that it is designed to investigate changes in protein levels. Therefore, it is particularly well suited for studying processes where proteins are likely to change, such as those involving the ubiquitin-proteosome protein degradation pathway (Chen et al., 2012). Proteomics was used to identify proteins modified by ubiquitin (Kim et al., 2011). Label-free proteomics was been used to investigate the proteins that are differentially degraded in human cancer cells lacking a specific E3 ligase in order to identify

its targets (Burande et al., 2009). However, it is clear that proteomics is a global analysis tool that can only generate hypotheses that need specific experimentation in order to verify them.

1.9. Aims of the project and organisation of the thesis.

The purpose of the Introduction of Chapter 1 was to introduce the idea that modification of Golgi function has effects beyond vesicular transport. It is intimately tied to the overall functioning of the cells through a number of signalling systems. Ubiquitinylation was the focus of this introduction, because manipulation of Ub processes is one of the proposed functions of the Ftp105/Hid3 protein of *S. pombe*. However, it is likely that Ftp105/Hid3 has other function in the Golgi or can affect other processes when mutated. There is also the question of what are the two Ftp105/Hid3 paralogues doing. The group work, which this thesis project is part of, is using post-genomic technologies of transcriptomics, proteomics, and metabolomics to give more clues as to what these proteins are doing., to investigate the stress responses of mutants without the *hid* genes replaced, and then to determine the consequences on the metabolome and proteome of the mutants.

Chapter 2 presents the methodology relating to the experimentation described in this document.

Chapter 3 describes the work to computationally analyse the HIDs and their relationships to their closest members the DYMs. Animals have both HID and DYM and their localization and action, such as dynamic association with the Golgi, appear to be very similar. The phylogenetic work was conducted to determine if there is any evidence that the two proteins have overlapping functions. It is evident that eukaryotes can have either DYM or HID or both, but animals have both. Fungi, except for rare exceptions, only have HID and plants only have DYM. The study confirmed the presence of three HID1 orthologues in *S. pombe* and presented a sequence of diversification. I suggest that Ftp105 be called Hid3 in the future. Lastly, publically available data was mined in order to find potential associations of HID with other proteins that might reveal aspects of function.

Chapter 4 describes the testing of the *hid* mutants under different types of stresses. Previously published data suggested that *hid1+* and *hid2+* may be involved in stress responses, because they are induced in response to stress (Chen et al., 2003). The mutants analysed were produce in the study of Alshehri (2015). All mutants *hid1Δ*, *hid2Δ* and *hid3Δ* were insensitive to

a wide variety of stresses, except oxidative stress that affected both of *hid1Δ* and *hid3Δ* by greatly slowing growth of cultures. Interestingly, the growth of *hid3Δ* was severely affected when

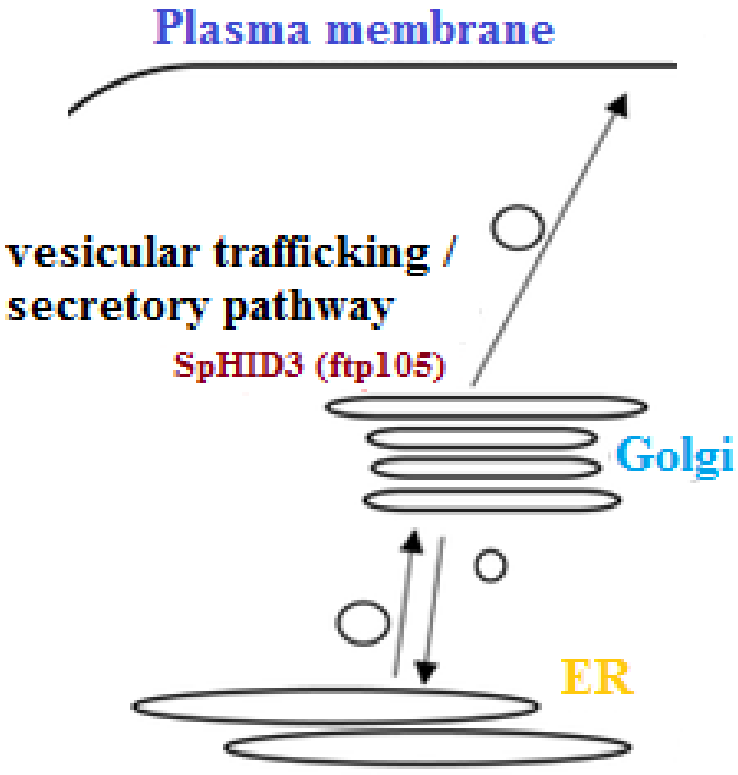


Figure 1.4. Schematic diagram of the possible role of *S. pombe* Ftp105/Hid3-mediated protein trafficking from the Golgi.

grown on minimal media. Why this was the case is unknown, but I speculate that *hid3Δ* is lacking an essential nutrient. Lastly, by use of the Golgi inhibitors brefeldin A and golgicide A, I was able to determine that Hid3 in *S. pombe* functions in anterograde protein transport (Figure 1.4).

Chapter 5 describes the metabolomics and proteomic analysis of the mutants. Metabolomics is used to judge how greatly the fundamental physiology and biochemistry of cells change when a genetic mutation is introduced. Also, metabolic changes can frequently explain why cellular growth is modified. The purpose for conducting a proteomic study directly relates to the potential involvement of Hid3 in ubiquitin-mediated processes that would involve protein degradation. It is clear that proteomics is broad-scope and there can be a number of ways that affecting Golgi may change protein levels, but the first step was to determine which proteins change and by how much. Both the proteomics and the metabolomics were quite revealing in that they clearly show *hid3Δ* cells responding as though they were stressed with up-regulation of both stress signalling proteins and oxidative stress response proteins and down-regulation of protein synthesis genes.

Chapter 6 is the general discussion that gives the biological context of the results in terms of stress responses and metabolite and protein changes. I discuss how the results may support the idea that slow growth may be a cause of cells in a partial state of quiescence. I also present some detailed mechanisms through the protein data showing how disrupted Golgi function can alter chromatin structure and affect gene expression.

Chapter 2

Materials and Methods

2.1 Phylogenetic and computational analyses

Ortholgous and paralogous protein sequences of the HIDs and DYMs were identified by DELTA-BLAST analysis of the non-redundant protein database at NCBI. The full-length sequences matching the cut-off criteria were obtained from NCBI. All alignment and phylogenetic steps were conducted using MEGA6 (Tamura et al., 2013). MUSCLE (Edgar, 2004) was the primary alignment algorithm used. Tree construction was done in MEGA6 by either the Neighbour Joining or Maximum-Likelihood methods. Confirmation of functional identity as either DYM or HID, that is no other conflicting identification, was done by re-screening of the NCBI, or a database for a specific organism (e.g. Wormbase, Pombase, etc.) by BLAST analysis. Closer phylogenetic relationships were investigated through syntenic relationships among organisms, such as the fungal HIDs through the YGOB genome browser, and protein structural and functional features found in ProDom, pFam, InterPro and UniProt. Mapping of protein-protein interactions was performed using the Strings database (www.string-db.org).

2.2 Charaterization of mutant growth and stress responses

2.2.1 Growth of *S. pombe*

All strains of *S. pombe* were stored at -80 ^{o}C in YES medium (0.5 % w/v yeast extract, 3 % w/v glucose, 225 mg/l each of adenine, leucine, uracil, histidine and lysine) containing 25 % reagent grade glycerol (VWR). Cells were grown on YES/Agar (1.4 %, w/v) plates at 30 ^{o}C for 3 days for short-term storage at 4 ^{o}C. Routine growth of cells was in 5 ml liquid cultures of YES media at 30 ^{o}C with shaking at 200 rpm. Edinburgh Minimal Medium (EMM) and EMM with agar adenine, histidine and lysine were purchased from Sunrise Science Products via ELIS Solution (Erpent, Belgium). In order to test for growth defects under normal conditions, wild-type and mutant strains were inoculated into 5 ml of YES medium and incubated for 12 hours at 30 ^{o}C with shaking at 200 rpm. Cells were diluted into fresh YES or EMM medium at a cell number of 5 x 10^5 cells/ml and incubated at 30°C with shaking at 200 rpm.

Table 2.1 Chemical Stress agents. The following table provides the type and concentration of the chemical stress treatment and the solvent in which it was reconstituted.

Drug	Vendor	[Working]	[Stock]	Solvent
Hydroxyurea (H.U)	Sigma	3mM, 5mM and 10mM	0.1g	10 ml DMSO
Methyl Methanesulfonate	Fluka	0.0025%, 005% and 0.0075%	100µl	9.9 ml H_2O
Thiabendazole (TBZ)	Sigma	10, 12.5, 15, 20, and 25 µg/ml	0.1g	10 ml DMSO
Camptothecin	Sigma	0.5 µg/ml and 0.8 µg/ml	0.1g	10 ml DMSO
Phleomycin	InvivoGen	1, 2.5, 5 and 10 µg/ml	0.1g	10 ml H_2O
Brefeldin a	Sigma	5, 15, 20, 25 µg/ml	10mg	1 ml DMSO
Golgicide a	Sigma	50 µg/ml	10mg	1 ml DMSO

Table 2.2. List of genotypes used for experimentation. All mutant genotypes include the parental, wild-type genotype as well as the modified *hid* locus.

Strain	Genotype	Source
Wild-type	*ade6-m26 ura4-D18 leu1-32*	McFarlane, BP90
VCK	*kanMX6*	Alsherhi 2015
VCN1	*natMX6*	Alsherhi 2015
VCN2	*natMX6*	Alsherhi 2015
VCU1	*ura4*$^{+}$	Alsherhi 2015
hid1Δ iso1	*hid1Δ::natMX6*	Alsherhi 2015
hid1Δ iso3	*hid1Δ::natMX6*	Alsherhi 2015
hid2Δ iso1	*hid2Δ::ura4*$^{+}$	Alsherhi 2015
hid2Δ iso2	*hid2Δ::kanMX6*	Alsherhi 2015
hid2Δ iso3	*hid2Δ::kanMX6*	Alsherhi 2015
hid3Δ iso1	*hid3Δ::natMX6*	Alsherhi 2015
hid3Δ iso2	*hid3Δ::natMX6*	Alsherhi 2015
hid3Δ iso3	*hid3Δ::kanMX6*	Alsherhi 2015
hid3Δ iso4	*hid3Δ::kanMX6*	Alsherhi 2015
hid3Δ iso3 *hid2Δ* iso1	*hid3Δ::kanMX6 hid2Δ::natMX6*	Alsherhi 2015
hid3Δ iso3 *hid2Δ* iso2	*hid3Δ::kanMX6 hid2Δ::natMX6*	Alsherhi 2015
hid3Δ iso4 *hid1Δ* iso1	*hid3Δ::kanMX6 hid1Δ::natMX6*	Alsherhi 2015
hid3Δ iso4 *hid1Δ* iso2	*hid3Δ::kanMX6 hid1Δ::natMX6*	Alsherhi 2015
hid2Δ iso1 *hid1Δ* iso1	*hid2Δ::ura4*$^{+}$ *hid1Δ::natMX6*	Alsherhi 2015
hid2Δ iso1 *hid1Δ* iso2	*hid2Δ::ura4*$^{+}$ *hid1Δ::natMX6*	Alsherhi 2015

2.2.2 Growth analyses of *hid* mutants under stress conditions

The creation of gene replacement strains from the haploid wild-type strain (*h*$^{-}$ *ade6-m26 ura4-D18 leu1-32)* was done according to Bähler et al. (1998) and the detailed procedure and molecular characterisation of the mutants is given in Alshehri (2015). All strains were grown for ~12 h in 5 ml YES at 30^{o}C with shaking at 200 rpm and diluted to a cell count of 5 x 10^{5} cells/ml into 5 ml of fresh media containing a the stress-inducing chemical agent. The chemical agents used, the vendor, the working and stock concentrations used and the solvent in which the agent was dissolved are given in Table 2.1. For abiotic stresses, such as heat or cold, diluted cultures were placed under appropriate conditions with shaking at 200 rpm. Cold and heat treatment consisted of growth of cells at 23 ^{o}C and 37 ^{o}C, respectively. All stress treatments lasted for 12 hours. The genotypes subjected to stress treatments are given in Table 2.2. Growth data is generally presented as Generation Time (GE), i.e. the doubling-time of the culture, and was calculated from starting and end cell counts according to the equation GE =12/(3.3*Log($N_{t=12}$/$N_{t=0}$) where $N_{t=12}$ was the number of cells at 12 h and $N_{t=0}$ was the starting number of cells. S.pombe cells were grown on YES medium appropriate supplements. To block ER-to-Golgi transport, cells were treated with various concentrations of BFA 0, 5 and 25 µg/ml for 12 hours. Cells treated with DMSO were used as controls. A concentration of 50 µg/ml was used for GCA to test the strains.

2.2.3 Determination of cell number

Three different methods were employed to determine the cell number within cultures for the growth analyses. Rough estimates of cell number and relative growth rates were obtained through spectrophotometric analysis of cultures taken at a wavelength of 600 nm. Precise numbers of cells were obtained by counting under the microscope using a Thoma haemocytometer (Life Technologies, catalogue number W30157544). Routine counting of cells for the growth analyses was done using a Nexcelom Cellometer® Mini automated cell counter and 20 µl chamber counting slides (Ozyme Inc.). For each growth experiment, the overnight culture was counted and diluted accordingly and the dilution culture recounted to obtain the starting cell number.

2.3 Genotyping of strains of *S. pombe*

Since many different mutants strains were used in the analysis of growth properties, strains were genotyped by PCR routinely to verify their authenticity.

2.3.1 Isolation of the DNA from *S.pombe* deletions mutants

The isolation of DNA was as follows. All strains were grown in liquid cultures of 5 ml of YES to late log-phase, and 3 ml of cells were harvested by consecutive centrifugation in a microcentrifuge spun at 6,000 rpm for 5 mins. The supernatant was removed and the pellet was washed in 1 ml of sterile distilled H_2O containing 0.1% (w/v) Sodium azide. DNA was isolated from cell using the Wizard™ Genomic DNA isolation kit according to manufacturer's instructions (Promega). In brief, the solution was centrifuged at 6,000 rpm for 2 min to pellet the cells and the supernatant was removed. The cells were resuspended thoroughly in 300µl of a buffer containing 50mM Na_2HPO_4, 11.5g/l citric acid, 40mM EDTA, 1 M sorbitol, and 0.1 mg/ml Zymolyase® 20T (Amsbio), and the tubes incubated at 37°C for 30 min to digest the cell wall. After cooling the tubes to room temperature (~ 2 min), the cells were again pelleted by centrifugation in microfuge at 13,000rpm for 2 min, and the supernatant discarded. Next, 300µl of Nuclei Lysis Solution was added to the tubes, the pellets were resuspended by gentle mixing and 100µl of Protein Precipitation Solution was added. The tubes were vortexed vigorously for 20 seconds and the tubes incubated on ice for 5 min. After centrifugation for 2 min and the supernatant containing the DNA was transferred to a clean 1.5ml microcentrifuge tube containing 300µl of isopropanol at ambient temperature and gently mixed by inverting the tubes until the DNA precipitate became visible. The DNA was collected by centrifugation, the supernatant was removed, and the pellet was washed with 300µl of room temperature 70% ethanol by inverting the tubes several times. The DNA pellet was recollected by centrifugation for 2 min, the ethanol was discarded, and the pellet was allowed to dry by leaving the tube inverted on clean absorbent paper for at least 10 min. The DNA was dissolved in 50µl of DNA Rehydration Solution to which was added 1.5µl of a 1 mg/ml RNase Solution and the tubes incubated at 37°C for 15 minutes. Finally, the DNA was left to rehydrate overnight at 4°C.

2.3.2 Polymerase chain reaction

The primer pairs used to genotype mutant strains were 1) external to the site of insertion, 2) two sets of insert/genomic junction primers and 3) a set of internal primers for each of the marker and *Hid* genes. All primers were designed using Primer3 (v 0.4.0, http://bioinfo.ut.ee/primer3-0.4.0/) and the corresponding Sp*Hid* gene sequences as templates. All PCR primers were obtained from eurofins Genomics (http://www.eurofinsgenomics.eu).

All primers were tested by PCR on appropriate DNA or cDNA templates. PCR reactions were prepared in a volume of 10 µl consisting of 1x GoTaq Flexi Buffer, 3 mM $MgCl_2$, 0.2 mM dNTP, 0.05 µM of forward and reverse primers, 0.25 Units of GoTaq DNA Polymerase, 10 ng of

DNA. PCR was performed on an MJ Research thermocycler with a sequence of steps comprised of an initial denaturation step of one cycle of 1 min at 94 °C, 34 cycles of 15 sec at 94 ^{o}C, 15 sec at 55 ^{o}C, and 15 sec at 72°C, followed by an extension period of 72ºC for 5 min. PCR products were observed by agarose (1 % w/v) gel electrophoresis using SafeView® (NBS Biologicals) as detection system.

2.4 Proteomic analyses

The strategy for combining the various mutant genotypes to yield 3 biological replicates for wild-type, negative control and *hid1Δ* and *hid3Δ* strains is given in Chapter 5.

2.4.1 Growth of cells

Starter cultures of 5 ml of cells for wild-type, negative control and mutant strains were grown ~ 12 hours in 5 ml volumes of YES at 30 ^{o}C with shaking at 200 rpm. The starter cultures were used to inoculate 50 ml volumes of YES to a cell count of 5 x 10^5 cells/ml, which were again incubated at 30 ^{o}C until each culture had reached a final cell count of 5 x 10^7 cells/ml. The timing of incubation was adjusted according to growth rate in order for each culture to reach this concentration of cells. Cells were harvested by centrifugation for 5 min at 6,000 rpm at 4 ^{o}C and immediately frozen in liquid N_2. Frozen cell samples were combined at the step of homogenization to yield three independent biological replicates each containing a mixture of two mutant or control genotypes. Three independent cultures were combined to give each biological replicate of wild-type.

2.4.2 Extraction of total cellular protein

Total protein was extracted from a quantity of 100 mg of frozen, powdered cells (FreezerMill 6750) as specified in the Fission Yeast Handbook. The powdered cells were added to 2 ml BeadBug™ prefilled tubes (Sigma-Aldrich, Z763748), which had been cooled in liquid N_2. A volume of 500 µl of HB buffer composed of 25 mM MOPS pH 7.2, 60 mM ß-glycerophosphate, 15 mM p-nitrophenylphosphate, 15 mM MgC12, 15 mM EGTA, 1 mM DTT, 0.1 mM sodium vanadate, 1% Triton X-100, 1 mM PMSF, 1 x Yeast Protein inhibitor cocktail (Sigma-Aldrich P8215) and 1 x mammalian protease inhibitor cocktail (Sigma-Aldrich, P8340) was added to the tube. Just as the mixture was beginning to thaw, the cells were broken open by shaking the tubes for 2 x 30 secs at a Frequency (1/s) of 30 in a TissueLyser II (Qiagen). The

tubes were cooled for 30 s on ice between each shaking. The tubes were centrifuged at 13000 rpm at 4°C for 5 mins and the supernatant transferred to clean tubes for storage at -20°C.

2.4.3 Protein quantification and visualization

The amount of total protein in extracts was determined using the Pierce BCA™ Protein Assay Kit according to manufacturer's instructions (Perbio) and verified using the method of Bradford with the reagent purchased from Sigma-Aldrich. The BSA included within the BCA kit was used routinely to produce standard curves for quantification for both methods. The volumes of solutions for both methods were adjusted proportionally to those specified in the instructions to permit coloration reactions to be done directly in 1.5 ml microfuge tubes. Visualization of total protein extracts was done by discontinuous 1D-SDS-PAGE using 5% and 10% w/v bis-acrylamide for stacking and separating gels, respectively (Laemmli, 1970). The pHs of the stacking and separating gels were 6.8 and 8.8, respectively. The running buffer was 1 x TRIS-glycine-SDS prepared according to (Sambrook et al., 1989). Laemmli Sample Loading Buffer was purchased from Bio-Rad (cat # 1610747) to which 2-mercaptoethanol was added prior to preparation of samples. PageRuler™ unstained and pre-stained markers were purchased from Thermo Scientific Ltd. All PAGE was conducted using a MiniProtean™ II unit (Bio-Rad). Proteins were visualised in-gel using QC colloidal commassie according to manufacturer's instructions (Bio-Rad).

2.4.4 Label-free proteomic analyses

Ten µg of total protein from each sample was loaded onto a 10% SDS-PAGE and allowed to migrate until all protein had just entered into the gel. Proteins were visualized by staining with colloidal blue (Biorad) and from each lane, a gel slice containing the protein was excised. Proteins were digested in gel using trypsin and the peptides eluted from the gel. The samples were acidified and injected into a Q-Exactive™ quadripole Orbitrap mass spectrometer (Thermo Scientific) and the data collected over a liquid chromatography run period of 120 min. The resulting mass spectra were interrogated with Proteome Discoverer™ v 1.4 (Thermo Scientific) using the *S. pombe* proteome Discoverer™ v 1.4 software as template for identifying peptide masses. Proteins changing significantly among mutant and control genotypes were determined by ANOVA (threshold, $p < 0.05$) and only proteins with a minimum of two peptides were quantified for differential amounts. PCA was conducted in R using the work package FactoMiner. For the GO analysis, I used all proteins identified by MS data (2107 proteins) and then I used UniProt batch retrieval to convert protein names to transcript names. That resulted

in total of 2309 genes. The changed genes files were constructed by taking those proteins that were 1.5-times up or down-regulated in each genotype *hid1Δ* and *hid3Δ* as compared to the negative control as explained above. Data for both genotypes were submitted together to GoMiner web server (http://discover.nci.nih.gov/gominer/GoCommandWebInterface.jsp). The settings used for the analysis were data course set to Pombase.org with the organism being *S. pombe*, a bootstrap analysis with 100 randomisations of the data and the gene class size was set to range from 5-500. The analyses for the each catergory Biological Process, Cellular Component and Molecular Function were done individually.

2.5 Immunological analyses of Hid proteins

2.5.1. Production of Antibodies

Antibodies were commercially produced in rabbit using synthetically generated peptides by Perbio (Thermo Scientific). The peptide sequences were designed using translation sequences of each *Hid* gene taken from Pombase taking into account the diversity of sequences to minimize cross-reactivity. The peptide sequences are given in Chapter 5. The immunization protocol was done over a 90 day period entailing a primary injection and 2 booster injections with Freund's adjuvant. Sera were provided from pre-immune bleeds and after first and second boosters. Two rabbits were immunized with each peptide and the rabbit with the best titer after the first booster was selected for terminal bleed. Two subsequent booster immunizations were used to increase the titer of anti-SpHid1 and anti-SpHid2 antibodies. A portion of the antibodies from the terminal bleed were purified and the purified antibodies used for testing recognition of the various Hid proteins.

2.5.2 Immunoblotting

Protein and peptide were loaded directly onto PVDF membrane using a slot-blot device. The visual loading of protein was done using using 0.01% Ponseau S. Immunoblotting was conducted using the WesternDot™ 625 Western Blot Kit (Invitrogen, Life Technologies). The purified anti-Hid antibodies were used for immunoblotting at a dilution of 1/500 in the appropriate incubation buffer provided with the kit. Most incubations were of first antibody were allowed to proceed for at least 12 hours. Incubations with secondary anti-rabbit antibodies were allowed to proceed for at least 2 hours. After development according to the manufacturer's instructions, the images of blots were taken using a Bio-Rad EZ-imager™.

2.6 Metabolite profiling

Metabolites were extracted from approximately 30 mg of freeze-dried *S. pombe* cells grown as described in Section 2.1.2.1. Metabolites were extracted according to (Weckwerth et al., 2004)) and quantified by ^{1}H-NMR as described in (A et al., 2004). Briefly, the cells were extracted with a mixture of H_20/ethanol/chloroform and the aqueous phase recovered and dried under vacuum. The extracts were resuspended in 200 mM phosphate buffer pH 6.0 in D_2O and analyzed at 500.162 MHz on a Bruker spectrometer (Bruker Biospin Avance). Metabolites were quantified using the metabolite mode of AMIX software (Bruker Biospin v. 3.5.6) based on the number of protons comprising the corresponding resonance. Concentrations in the NMR tube were converted to amounts per g fresh weight using the mass of sample extracted. Citrate, formate, fumarate, glutamate, lactate and malate are expressed as µg of the acid form. The concentration of NMR unknown compounds (named according to the form of the resonance, S for singlet, D for doublet, M for multiplet, and its frequency in ppm) was calculated on the assumption that the measured resonance corresponded to one proton and using an arbitrary molecular weight of 100 Da.

2.7 General Data Analysis.

Principal Component Analysis and other statistical analyses were performed in R using various work packages. Specific aspects of data analysis and how they were performed are described in the appropriate chapters.

Chapter 3

Computation analyses of the DYMECLIN protein superfamily and the HID-1 orthologues from *S. pombe*

3.1 INTRODUCTION

Computing has become an essential aspect of biological research. The production of large data sets has driven an interest to understand the interaction of biological components (Wooley et al., 2005). The start of any research project should begin with a detailed investigation about what is currently known on a topic, not only through literature searches, but extracting relevant information from the large number of databases. In addition, the advancement of computing has enabled investigations into aspects of biology not previously possible. From sequences of DNA we can now predict protein sequences and patterns that permit in-depth comparisons of protein function. Through phylogenetic analyses we can observe how protein function varies among organisms and also allows us to estimate the evolutionary origin of particular genes and proteins. This chapter presents important background taken from public sources that has been used as a guide for my experimentation and that of the group in general.

As presented in the general introduction of Chapter 1, the term HID, used to describe the gene and protein, was introduced in a report describing the isolation of constitutive dauer phenotype in mutants of *C. elegans* (Ailion and Thomas, 2003). The human ortholog was, until just recently, called either DMC1 for Down-Regulated in Multiple Cancers 1 (Harada et al., 2001) or C17orf28 by the nomenclature for uncharacterized genes from the human genome sequencing and annotation project. Therefore, the function of HID1 in humans is not well characterized, probably because it does not have as severe a phenotype as DYM when mutated. It appears to be more critical in the stress responses of invertebrates through the formation of Dauer larvae. It is interesting to speculate that DYM is not a lethal mutation in humans, because it could be partially complemented by HID1.

The aim of this chapter is to present currently available information resident within database to show that S.pombe is a suitable model system for investigating HID function. The computational aspect starts with a phylogenetic analysis of the relationship of the HIDs and DYMs as they are both part of the DYMECLIN superfamily. The purpose of the phylogenetic study is not to determine the accuracy of current ideas of taxonomic relationships, but simply to

investigate if HID and DYM share a common evolutionary path. This may help to show if either one or both are necessary for fundamental cell function. Therefore, we find it appropriate to base our sequence sampling strategy on the taxonomic relationships presented in NCBI that have tried to include various models and not proposed a single evolutionary model. A more in-depth comparison of protein properties is presented to determine common structural features among the S. pombe and animal HIDs to try to determine why Ftp105 appears to be the functional homologue of the animal HIDs. Lastly, an investigation into potential interactions of the S. pombe Hids was conducted to develop more their biological functions to guide further work.

3.2 RESULTS

3.2.1 Global phylogenetic analysis of HIDs and DYMs

It is interesting that pfam describes the HID motif as conserved from fungi to animals and that for the DYM as conserved from plants to animals. This would imply that 1) the proteins are only of eukaryotic origin and 2) that one or more of the major classes of eukaryotic organisms do not possess either a HID or a DYM, or both. Although a comprehensive phylogenetic and functional analysis of the HIDs represents a major work on its own, it was possible to use protein sequences present in NCBI in order to address the questions asked above. Also, it was possible to determine if the two protein families have similar evolutionary paths.

3.2.1.1 Selection of sequences from eukaryotic taxa for phylogenetic analysis

For the global phylogenetic analysis of the HIDs, we queried the Genbank non-redundant protein sequences using the HIDs and DYMs from *Homo sapiens*, the HID-1 orthologue from *S. pombe* and DYM from *Arabidopsis thaliana*. DELTA-BLAST with default parameters was used except that the number of sequences retrieved was set at 5000 and the significance cut-off was set at 0.01. We selected those sequences that had a significance score below the cut-off, but also for which the sequence coverage was greater than or equal to 40%. We chose this coverage cut-off value, because we would be likely to pull-out sequences from both families using a single query sequence. For example, DELTA-BLAST with AqHID or HsHID retrieves two proteins from each of the two choanoflagellates, *monosiga brevicolis* and *Salpingoeca rosetta*, and a 40% coverage cut-off separated them into their respective families. We also found that a 40% cut-off eliminated the vast majority of partial sequences. As described below, we identified a core sequence for the DYMECLIN superfamily that is around 200 aa, which can comprise up to 30% of the predicted full-length protein. Identification of a true core

Table 3.1. List of evolutionary lineages containing putative DYMs or HIDs. The HIDS from *H. sapiens* and *S. pombe* and the DYMS from *H. sapiens* and *A. queenslandica* were used to screen NCBI Genbank for putative homologues at the phylum or subphylum level. Sequences were recovered that exhibited a DELTA-BLAST score of <0.001 and sequence coverage of at least 40%. The highest-level scientific names were taken from NCBI taxonomy and their classification as given in Adl et al. (2005) or Cavalier-Smith (1981, 1998).

Evolutionary classification	HID	DYM	Genome Sequence
Alveolata (First Rank)			
Acavomonas	N	N	N
Apicomplexa	Y	Y	Y
Chromerida	N	N	Y
Ciliophora	Y	N	Y*
Colpodellidae	N	N	N
Colponemidia	N	N	N
Dinophyceae	N	N	Y*
Ellobiopsidae	N	N	N
Perkinsea	Y	N	Y
Veromonas	N	N	N
Unclassified	N	N	N
Amoebozoa (Super-group)			
Acromoeba	N	N	N
Archamoebae	N	N	Y*
Discosea	Y	Y	Y
Gracilipodida	nd	nd	N
Multicilia	nd	nd	N
Mycetozoa	Y	Y	Y
Phalansterium	N	N	N
Stereomyxa	N	N	N
Telaepolella	N	N	N
Tubulinea	N	N	Y*
Unclassified	N	N	N
Apusozoa (Phylum)	N	N	Y*
Breviatea	N	N	N
Centroheliozoa	N	N	N
Cryptophyta (First Rank)			
Cryptomonadales	N	N	Y*
Pyrenomondales	Y	Y	Y
Urgorri	N	N	N
Unclassified	N	N	N
Euglenozoa (First Rank)			
Diplonemida	N	N	N
Euglenida	N	N	Y*
Kinetoplastida	Y	Y	Y
Symbiontida	N	N	N
Fornicata (First Rank)	N	N	Y*
Glaucocystophyceae (First Rank)	N	N	Y*
Haptophyceae (First Rank)			
Coccolithales	N	N	N
Coccosphaerales	N	N	N
Isochrysidales	Y	Y	Y
Pavlovales	N	N	Y*

Phaeocystales	N	N	Y*
Prymnesiales	Y	Y	N
Reticulosphaerales	N	N	N
Syracosphaerales	N	N	N
Zygodiscales	N	N	N
Unclassified	N	N	N
Heterolobosea (First Rank)			
Acrasida	N	N	Y*
Creneidae	N	N	N
Pharyngomonas	N	N	N
Psalteriomonadidae	N	N	N
Schizopyrenida	Y	Y	Y
Tulamoebidae	N	N	N
Unclassified	N	N	N
Jakobida (First Rank)	N	N	Y*
Katablepharidophyta (Class)	N	N	N
Malawimonadidae (First rank)	N	N	Y*
Opisthokonta (Super-group)			
Choanoflagellida	Y	Y	Y
Fungi	Y	Y	Y
Metazoa	Y	Y	Y
Nucleariidae	N	N	Y*
Incertae sedia	Y	Y	Y
Oxymonadida (Second rank)	N	N	N
Parabasalia (First rank)			
Cristamonadida	N	N	N
Honigbergiellida	N	N	N
Hypermastigia	N	N	N
Hypotrichomonadida	N	N	N
Kofoidia	N	N	N
Spirotrichonymphida	N	N	N
Trichomonadida	N	Y	Y
Trichonymphida	N	N	N
Tritichomonadida	N	N	N
Unclassified	N	N	N
Incertae sedis	N	N	N
Rhizaria (Super-group)			
Acantharea	N	N	N
Cercozoa	N	N	Y*
Foraminifera	Y	N	Y
Gromiidae	N	N	N
Haplosporidia	N	N	N
Mikrocytiidae	N	N	N
Polycystinea	N	N	N
Sticholonchida	N	N	N
Unclassified	N	N	N
Rhodophyta (First rank)	N	N	Y*
Stramenopiles (First rank)			
Actinophryidae	N	N	N
Bacillariophyta	Y	Y	Y
Bicosoecida	N	N	Y*
Blastocystis	N	N	Y*
Bolidophyceae	N	N	N
Chrysomerophyceae	N	N	N
Chrysophyceae	N	N	Y*
Developayella	nd	nd	N
Dictyochophyceae	N	N	N

Eustigmatophyceae	Y	N	Y
Hyphochytriomycetes	N	N	Y*
Labyrinthulomycetes	N	N	Y*
Oikomonadaceae	nd	nd	N
Oomycetes	Y	Y	Y
Pelagophyceae	Y	N	Y
Phaeothamniophyceae	N	N	N
Pinguiophyceae	N	N	N
Placididea	nd	nd	N
PX clade	Y	N	Y
Raphidophyceae	N	N	Y*
Slopalinida	N	N	Y*
Synchromophyceae	N	N	N
Synurophyceae	N	N	Y*
Unclassified	N	N	N
Viridiplantae (Kingdom)			
Chlorophyta	Y	Y	Y
Streptophyta	N	Y	Y
Unclassified	N	N	N
Unclassified	N	N	Y*

nd No data retrievable from Genbank

structure for a targeted phylogenetic analysis based on domains was difficult at this time, because the length of this core sequence was highly variable and ranged between 70 and 400 AAs. However, a more in-depth investigation of this core structure may help to determine the minimal sequence comprising a HID or DYM motif and to unravel the evolutionary background of the DYMs.

The search was conducted using the 2nd-rank scientific names of NCBI Taxonomy in order to determine if any particular taxon possessed either a HID or a DYM or multiple copies of either. For the search of each taxon, sequences were selected according to the criteria of similarity mentioned above, and a 2000- iteration bootstrap of the Neighbor-joining algorithm was used 1) to pare-down the number of sequences to be included in the overall phylogenetic analysis, 2) to enable segregation of sequences into HIDs and DYMs, and 3) to reveal those species with multiple HIDs and/or DYMs. In general, the sequences selected were from species for which there was a genome sequence and from species with multiple HIDs or DYMs. Details on the choice of sequences selected are given for each scientific taxon searched. A list of the taxa searched and the presence or absence of HID or DYM homologues is given in Table 3.1.

For the Alveolata, the branch Apicocomplexa, which includes the genera plasmodium and toxoplasma, produced the most hits. Interestingly, the resulting phylogenetic tree showed a larger and a small subtree. The smaller subtree of 6 sequences grouped with the HIDs (primarily) and DYMs, whereas the larger group was further removed. The smaller group was comprised of the Toxoplasma, whereas the larger group contained species of Plasmodium and Cryptosporidium from the Ciliophora. The sequence from *Toxoplasma gondii* GT1 (gi|523573546|) was selected for the global phylogenetic analysis, since it was more similar to the query DYMs and HIDs. The more distantly related sequences from the species of the subphyla Ciliophora were not included in the global analysis. Two sequences from the Ciliophora were selected for the global analysis, one from *Oxytricha trifallax* (gi|403337977|) and one from *Paramecium tetraurelia* strain d4-2 (gi|145523772|). One sequence from *Perkinsus marinus* ATCC 50983 (gi|294952762|) represented the Perkinsea. Other subphyla within the *Alveolata* did not contain HID or DYM homologues.

For the Amoebozoa, sequences similar to DYM and HIDs were found only in the subphyla Mycetozoa and Discosea. The two branches were analysed independently to identify sequences for the global phylogenetic analysis analysis. From the Discosea, there were three sequences from *Acanthamoeba castellanii str. Neff*, one (gi|470387783) that groups with the

DYMs, another (gi|470454537) that groups with the HIDs and another (gi|470466316) that is more removed from the HIDs, but pulls up only HIDs with DELTA-BLAST. Each of the three sequences was included in the global phylogenetic analysis. For the Mycetozoa, the three proteins from *Dictostelium Fasciculatum* were also chosen, because they exhibit the same type of grouping as those from A. *castellanii*. One (gi|470244090|) was removed from both the DYMs and HIDs, whereas the two others, gi|470241913| and gi|470246996| grouped with the DYMs and HIDs, respectively.The search of the phylum Cryptophyta, yielded only two proteins, both of which were revealed with each of the four query sequences, and these come from *Guillardia theta*. This is the only member of the Cryptophyceae for which a genome sequence currently exists (http://genome.jgi.doe.gov/; (Curtis et al., 2012). Proteins gi|551677380| and gi|551670302| align with the HIDs and DYMs, respectively, thus these two sequences were included within the global phylogenetic analysis.

Searching of the Euglenozoa yielded numerous sequences similar to both DYM and HID of the sponge proteins. However, these sequences were only within the Order Trypanosomatida. Compilation of DELTA-BLAST significant hits using each of query proteins yielded a set of 42 sequences that were used for subsequent phylogenetic analysis. Species within five represented genera of the Trypanosomatida each contained two proteins, one each that clustered with the HIDs and DYMs. The one exception was *Trypsanosoma vivax*, which contained a third hypothetical protein similar to both DYMs and HIDS, and pulled up these sequences on a reciprocal BLAST query. Therefore, the three sequences from *T. vivax* (gi|340055718|, gi|340056098|, gi|340055981|) accessions were used for the global phylogenetic analysis. Reciprocal DELTA-BLAST with the remote sequence pulls up sequences of approximately 77 amino acids, which correspond to HID isoforms from fish species of the Order Cyprinodontiformes.

For the Haptophyceae, two sequences are pulled up from the genome sequence of *Emiliana huxleyi*, but only one of which appears to be full-length, gi|551625400|. MSA to the four query sequences shows the full-length protein to more closely resemble the DYMs than HIDs, with coverage and significance score highest with AtDYM. Only the putative full-length sequence was included within the global phylogenetic analysis as the partial sequence more closely resembled the paralogue than any of the potential orthologues.

For the Heterolobosea, the same two proteins from *Naegleria gruberi* were pulled up with each of the four query sequences. The protein gi|290972455| clusters with the HIDs and gi|291000556| aligns with the DYMs.

The Opisthokonta have both DYMs and HIDs in the metazoa and fungi, but which were restricted to specific orders. The taxa Incertae sedis of the Opisthokonts has the genus Capsaspora, also with both a DYM and a HID and others of unknown function. BLAST analysis of the fungi revealed two apparent paralogues of ftp105, which we term *SpHid2* and *SpHid3* according to divergence and we discuss these in more detail below.

Search of the Parabasalia pulls up only one protein from *Trichomonas vaginalis* G3 (gi|123457024|), which has greatest similarity to the DYMs.

Search of Rhizaria pulls up a protein from *Reticulomyxa filose*, which is significant, but only over a very short sequence of 26% query cover of the AtDYM with 37% identity. It has an identifiable DYM domain signature of 72 aa from aa 204 to 276. It is a marine protist and reciprocal blast pulls up HID homologues from primarily aquatic creatures through this domain. However, this domain also appears in terrestrial animals. DELTA-BLAST also recognizes another domain in subjects which corresponds to a short sequence > 70 aa starting at aa 420. This second domain is not recognized as a Dymeclin signature. Due to its short-sequence coverage, it was not included in the global phylogenetic analysis.

In order to determine which subfamily was present in the Viridiplantae, the Streptophyta and the Chlorophyta were searched independently using the each of the four query proteins. Each query pulled up the approximately the same number of significant hits, indicating that the search parameters were stringent enough to identify members of the other family. The union of the sequences retrieved were then aligned and subjected to phylogenetic analysis using Neighbor Joining and 2000 bootstrap replicates. The resulting phylogenetic tree showed that all Streptophyta proteins grouped with the DYMs. The proteins from *Oryza sativa*, *Arabidopsis thaliana*, and the lycophyte (Lycopsida) *Selaginella moellendorffii* were used in the global phylogenetic analysis. Inspection of the resulting Chlorophyta tree indicates that the sequences from the Trebouxiophyceae more closely resemble AtDYM whereas those from the Prasinophytes are more variable. The sequence from *Bathyococcus prasinos* (gi|612400476| and one from *Micromonas pusilla* (gi|003056438| are clearly related to the DYMs, but the sequences from the two Ostreococcus species were further removed from even AqDYM, and one sequences from Micromonas (gi|002504004| grouped with the HIDs. Therefore, the sequences selected for the global phylogenetic analysis were *Coccomyxa subellipsoidea C-169*

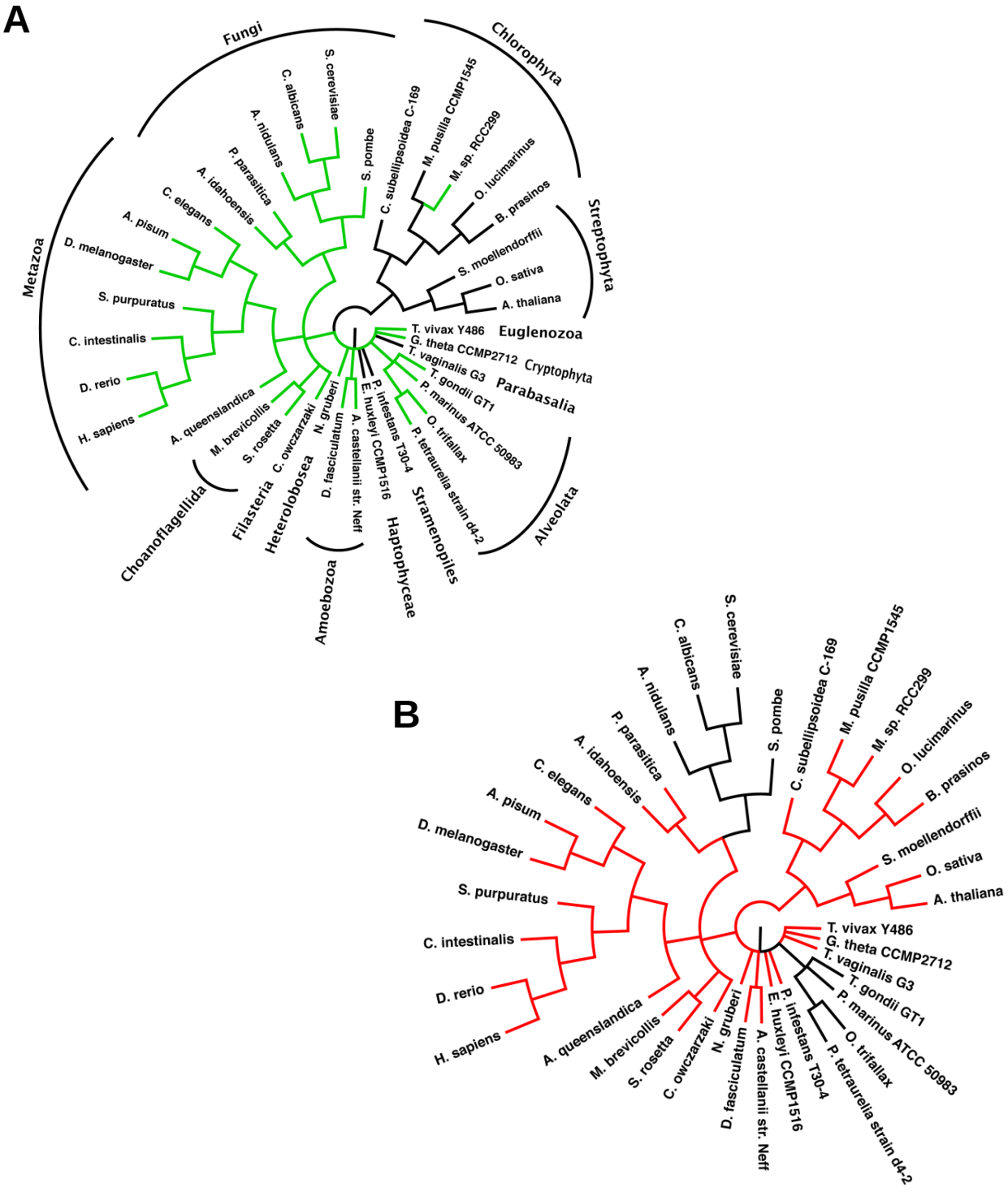


Figure 3.1. Presence of HID (A) or DYM (B) orthologues in the higher phylogenetic groupings. The branch topography represents the first two ranks of the Eukaryota according to NCBI taxonomy browser. Each phylogenetic group with either a HID or a DYM is represented by at least one species. Higher level ranks without HID or DYM are not shown. The coloured lines represent phylogenetic groups containing the respective protein.

(gi|545359709|) from the *Trebouxiophyceae* and all the Prasinophyte sequences, except *Ostreococcus tauri*, which was highly similar to *Ostreococcus lucimarinus* (gi|145341226|).

The list of the taxa searched and the presence or absence of HID or DYM homologues is given in Table 3.1. There were 110 scientific taxa according to NCBI Taxonomy searched. Of these 110, there were 82 taxa for which neither a HID nor a DYM was found and 26 of them are represented by a species with a sequenced genome. The next largest group contained 16 branches that had both HID and DYM. Six and two branches, respectively, had only HID or only DYM. The major branches that did not contain HID were the Stramenopiles, the Haptophyceae and the Viridiplantae (Figure 3.1A). The major branches that did not contain DYM were the Fungi, with two exceptions, and the Alveolata (Figure 3.1B).

Figure 3.2. Phylogenetic groupings of the HIDs and DYMs. (A) Neighbour-joining tree based on 2000 bootstrap replicates. (B) The bootstrap consensus tree of that shown in A. The trees were constructed in Mega6 using an alignment produced from MUSCLE (Tamura et al., 2013). The alignment was not manually modified prior to tree construction. The values represents the proportion of trees whereby removal of a sequence does not affect the overall topology of the tree. The lower the value the weaker is the support for the pairing. The scale bar represents the proportion of substitutions per site.

3.2.1.2 Global phylogenetic analysis

For the global phylogenetic, the set of protein sequences contained representative HIDs and DYMs from the BLAST searches of the individual taxa and the resulting phylogenetic tree construction. The set of proteins used contained 32 HIDs and 29 DYMs. Protein sequences from a variety of Metazoan species were used in order to determine if there was a consistent pattern of divergence between proteins within this particular taxon. This was only possible with the Metazoans noting that HIDs and DYMs were missing from the Viridiplantae and the Fungi, respectively. Phylogenetic trees were produced in Mega6 (Tamura et al., 2013) using both the Neighbour-Joining or Maximum-likelihood algorithms with at least 1000 bootstrap replicates. Both boot-strap and the boot-strap consensus trees from the Neighbour-Joining analysis is shown in Figure 3.2.

The overall structure of the trees shows two distinct families of proteins with the provisional designations correctly chosen. Starting from the top of the trees and the Metazoa, it was apparent that there is a continual divergence from the human HID-1 through the Metazoa with the *C. owczarzaki* and the choanoflagellate *M. brevicolis* and *S. rosetta* HIDs being the next most similar. Following on in divergence were HIDs from the other First-rank taxa, including the only HID from the viridiplantae, that from the Chlorophyta species *Micromonas sp. RCC299*. The fungal HIDs grouped together, except for ECM30 of *S. cerevisiae*, but were further diverged from the Metazoa than species from the other Supergroups. The most divergent HIDs included the ECM30 from *S. cerevisiae*, which paired with that from the Alveolata *P. tetraurelia*, and a second HID from the Amoeba *A. Castellanii* and a second HID from the Euglenozoan *T. vivax*.

The grouping of the Metazoa DYMs was similar to that for the HIDs with the species *C. owczarzaki* also part of the clade. Within this apparent clade were also the DYMs found within the Mucorales (fungal) species *P. parasitica* and *A. idahoensis*. Although not well supported by the bootstrap analysis, the adjoining clade contained four sequences, each from different first-rank taxa the Heterolobosea, Cryptophyta, Parabasalia and Amoebozoa. The next clade contained the DYMs from the Stramenopiles and Haptophyceae species. The Viridiplantae (Chlorophyta and Streptophyta) DYMs formed a separate clade one step further removed, whereas the furthest removed sequences were from the Amoeba *A. castellanii*, the Euglenozoa *T. vivax*, and lastly the sequences from the two choanoflagellate species.

The family of HID proteins contained species with apparent paralogues. Those most relevant to this work are the two paralogues of the *S. pombe* ftp105 and these will be discussed in more detail below. Not all paralogues grouped together for a species like they did in *S. pombe* and *P. parasitica*. The amoeba *A castellanii* has two apparent HIDs. One HID paired with that

from the other Amoeba species, *D. fasciculatum,* but the other HID paired with that from the Euglenozoan *T. vivax*. Furthermore, the *A. castellanii* DYM grouped with the Viridiplantae, whereas the DYM from *D. fasciculatem* paired with the DYM from the Heterolobosea *N. gruberi*. *T. vivax* also has two apparent HIDs. One HID is part of the diverse clade containing the Metazoa and other first-rank species. The second HID paired with *A. castellanii* as mentioned above. It was for these paralogues that the Maximum-the support for the pairing. The scale bar represents the proportion of substitutions per site.

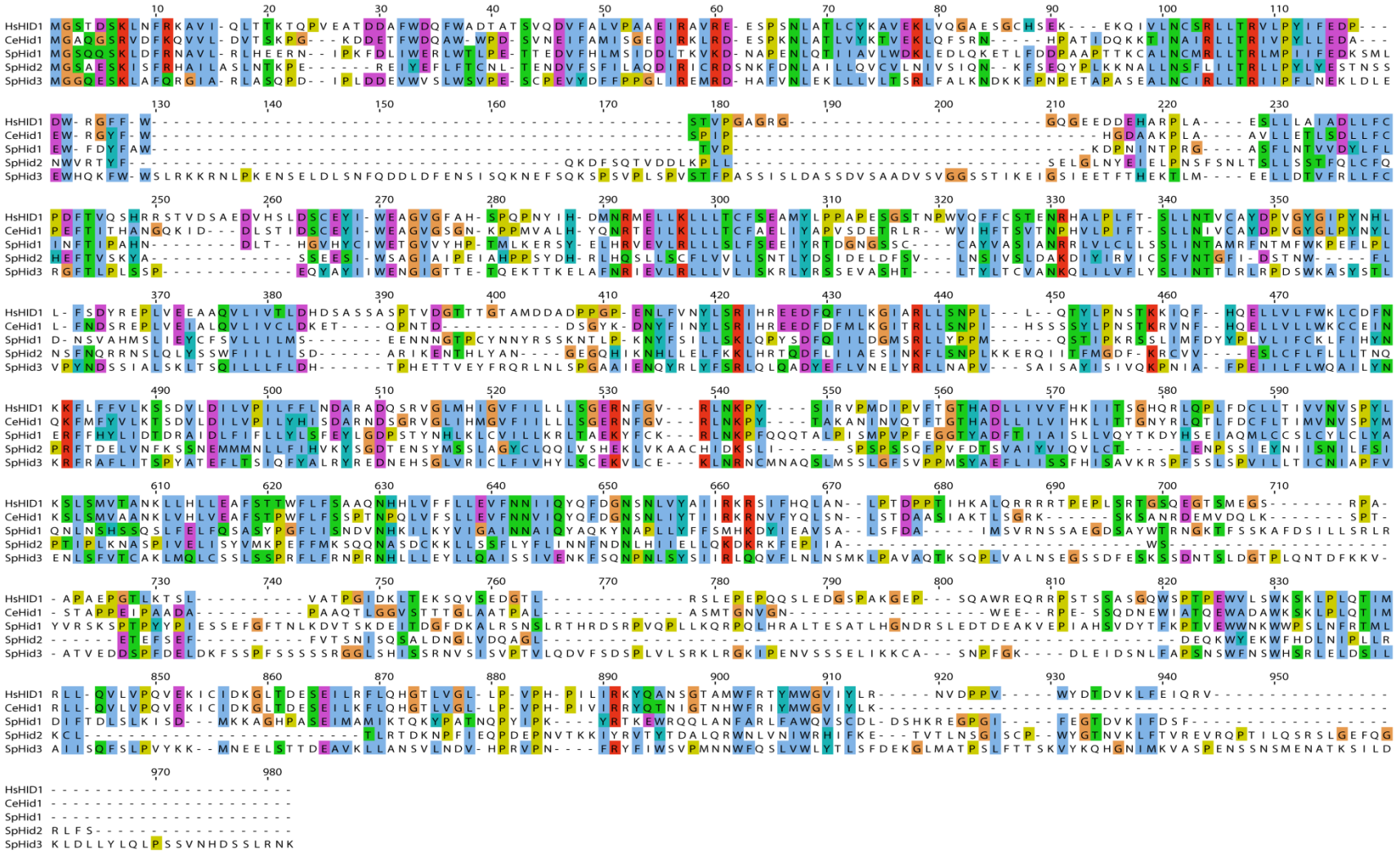


Figure 3.3. Multiple sequence alignment comparing *S. pombe Hids* with *HID1* from *C. elegans* and *H. sapiens*. The alignment was generated by MUSCLE in Mega6. Gaps, indicated by dashes, were incorporated to maximize the alignment, and positions of aligned residues are listed on the right side of the alignment. An asterisk indicates identical residues among the aligned sequences, a colon indicates conservation between amino acids with similar properties, and a period indicates conservation between amino acids of weakly similar properties.

Table 3.2. Quantitative comparison of primary sequences between members of the HID and DYM superfamily. The values for sequence identity and similarity were calculated using the Needle algorithm for pairwise sequence alignment from the EMBOSS suite of sequence analysis tools using the default settings (www.ebi.ac.uk; (McWilliam et al., 2013).

	HsHID1	CeHid1	DmHid1	ScECM30	SpHid1	SpHid2	SpHid3	HsDYM1	CeDYM1	DmDym1	AtDYM1
HsHID1		53.1 63.2	62.1 71.7	14.3 25.3	24.6 42.4	19.5 32.7	23.8 39.2	16.9 29.8	14.4 26.4	13.5 26.1	14.4 28.1
CeHid1			49.4 63.4	13.7 24.2	24.5 42.7	18.6 35.8	21.4 38.4	17.9 32.5	18.1 32.5	17.6 29.7	17.9 31.3
DmHid1				14.2 25.9	24.4 41.4	19.8 31.0	22.7 37.5	14.0 24.3	12.4 22.0	13.1 25.0	14.9 27.9
ScECM30					14.3 25.3	14.4 23.9	14.9 27.0	5.5 9.9	9.5 18.0	11.7 21.9	10.5 19.0
SpHid1						19.3 33.1	23.1 40.2	15.3 29.0	13.7 25.7	15.5 27.4	15.6 28.4
SpHid2							16.9 29.3	17.8 33.6	18.6 32.6	15.6 27.7	16.2 29.3
SpHid3								12.6 24.7	18.2 31.5	15.9 29.4	17.3 30.3
HsDYM									39.3 57.2	42.3 61.6	30.6 50.7
CeDym										32.9 54.5	23.6 40.9
DmDYM											27.9 45.2
AtDYM											

% Identity

- > 50 %
- 40-50 %
- 30-40 %
- 20-30 %
- 10-20 %
- < 10 %

likelihood tree differed from that of the Neighbour-joining one. It is expected that for the more diverse sequences where there is low bootstrap support for associations, that there would be differences among the various trees constructed.

A search for a core motif that could be used to facilitate alignment of sequences for the phylogenetic analysis, and which could be used to search early branching eukaryotic and bacterial species for potential ancestral proteins, has not been successful. This is because neither a definite HID nor a DYM protein family signature was clear. For the vast majority of sequences, the entire protein was designated as a HID or DYM motif. In pfam, DYM and HID1 have been defined as two families of the Golgi_traff clan (CL0456, pfam.xfam.org/clan/Golgi_traff) that contains 1159 domains representing 20 domain structures. Of the 754 DYM sequences, 664 have the DYM motif covering the entire sequence. Of the 405 HID1 sequences, 274 have the HID1 motif covering the entire sequence.

3.2.2 Functional characterisation of the *S. pombe* HIDs

3.2.2.1 Comparison of primary sequences to eukaryotic orthologues

The BLAST analysis of the Opisthokonta Super-group revealed that *S. pombe* contains three putative HID1 homologues. This was confirmation of the information already present in PomBase (www.pombase.org). The question then arose as if they were paralogues or if they were orthologous to other proteins. The large phylogenetic analysis suggests that they are paralogous due to similarity in sequence. At what point the paralogues came into existence could be addressed through an in-depth phylogenetic study of the Fungi, but such a study falls outside the scope of the chapter. Nevertheless, it is apparent that not all Fungi have 3 HIDs. The yeast *S. cerevisiae* appears to have one HID homologue ECM30 (Figure 3.2) and its sequence has greatly diverged from those of *S. pombe*. In order to determine the *S. pombe* HID that appears to be the direct orthologue of HID-1 from *C. elegans* and to establish the proper nomenclature for the *S. pombe* HIDs a multiple sequence alignment was conducted with the proteins HsHID-1, CeHID1, Ftp105, and the two Ftp105 paralogues, called hid1 and hid2 (Figure 3.3). The names of the two ftp105 orthologues were taken from PomBase. The pairwise identities and similarities are given in Table 3.2. SpHid1 and Ftp105 had nearly the same level of identity and similarity to CeHID1 and HsHID1 with just over 20% and 40%, respectively. SpHid2 was slightly more divergent with less than 20% identity and around 35% similarity. SpHid2 was almost as divergent from the other two S. pombe proteins as from those from humans and *C. elegans*. The pairwise comparisons with the DYMs from humans (HsDYM), C. elegans

Table 3.3. GO ontology references to potential function of Hid proteins from *S. pombe*. The gene ontology listings were taken from AmiGO 2 (amigo.geneontology.org).

Protein	Ontology listings	GO accession	Function	Ref.
Hid1	Nucleus	GO:0005634	cellular_component	1
	Cytosol	GO:0005829	cellular_component	1
Hid2	Nucleus	GO:0005634	cellular_component	1
	Cytosol	GO:0005829	cellular_component	1
Hid3	protein binding	GO:0005515	molecular_function	2
	Golgi apparatus	GO:0005794	cellular_component	1
	Cellular protein Localization	GO:0034613	biological_process	2
	Cytosol	GO:0005829	cellular_component	1

1 (Matsuyama et al., 2006)
2 (Kouranti et al., 2010)

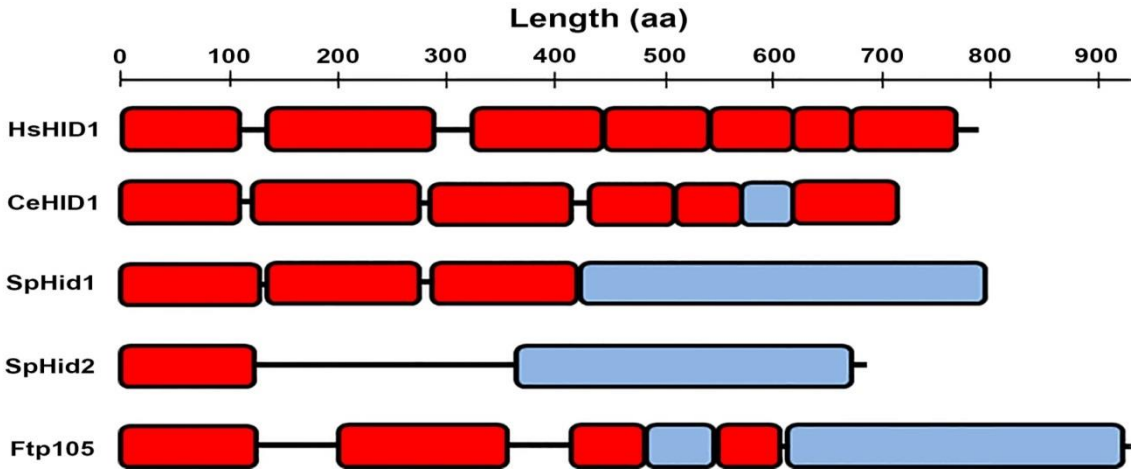


Figure 3.4. Comparison of HID domain structures. The determination and alignment of domains was done using the ProDom website. Searches for particular entries were done using the Uniprot designations: HsHID1, uniProt Q8IV36; CeHID1, uniProt Q7Z135; SpHid1, uniProt Q9P7M6; SpHid2, uniProt Q9HDV3; ftp105 uniProt O13776. The red rectangles represent transmembrane helix domains, the blue rectangles represent cytosol-related domains, and black lines represent undefined sequences. (CeDYM) and Arabidopsis (AtDYM) were also included to give a better idea of diversification than was obtainable from the trees shown in Figure 3.2. Interestingly, the degree of identity and similarity between the plant and animal DYMs was equal to or greater than those between the pombe and animals HIDs (Table 3.2).

3.2.2.2. Comparative structural properties of the HIDs from *S. pombe*

A search of the Gene Ontology Database AmiGO 2 revealed aspects of putative functions and/or roles for each of the *S. pombe* Hids (Table 3.3). From a genome-wide screen of protein localization, there is evidence that Hid1 and Hid2 can localize both to the cytosol and nucleus (Matsuyama et al., 2006), but no molecular or biological function has been given to these proteins. The same study put Hid3 into both the golgi and cytosol. There is some information about the molecular and biological functions of Hid3 that came from a study on the on the deubiquitinating family of enzymes in *S. pombe* (Kouranti et al., 2010). This study showed that this enzyme is involved in the transport of Ubp5 from the cytosol into the Golgi apparatus. Therefore, it has the molecular function of protein binding and the biological function of protein transport and localization.

Proteins are organised into structural units called domains that may have distinct functions (Alberts et al., 2002). Most eukaryotic proteins contain multiple domains that have a vital role in protein-protein interactions and interaction networks. Hid3 localization to the Golgi membrane and its involvement in the transport of Ubp5 from the cytosol into the Golgi would require it to have domains or features giving it protein and membrane interaction properties. Also, it is expected that the protein features of Hid1 and Hid2 will have diversified sufficiently from Hid3 in order to be able to detect particular differences among the proteins. These differences were tested by comparing domain features among HIDs from various organisms. The ProDom database (prodom.prabi.fr, (Servant et al., 2002) contains 56 different arrangements of domains covering 94 proteins. Included within this set is HsHID1, CeHID1, SpHid1, SpHid2 and ftp105. The structure of these proteins is mainly determined by the number and placement of the putative transmembrane helices (Figure 3.4). Prodom predicted seven transmembrane helices as has been previously reported (Wang et al., 2011). CeHID1 retained 6 of the corresponding transmembrane helices. The closest domain architecture of the S. pombe sequences to either of the animals was that for Hid1. In all three S. pombe Hids, the C-terminal helices have been replaced by long cytosolic domains. These C-terminal domains have sufficiently diversified that they are not considered to be similar according to ProDom.

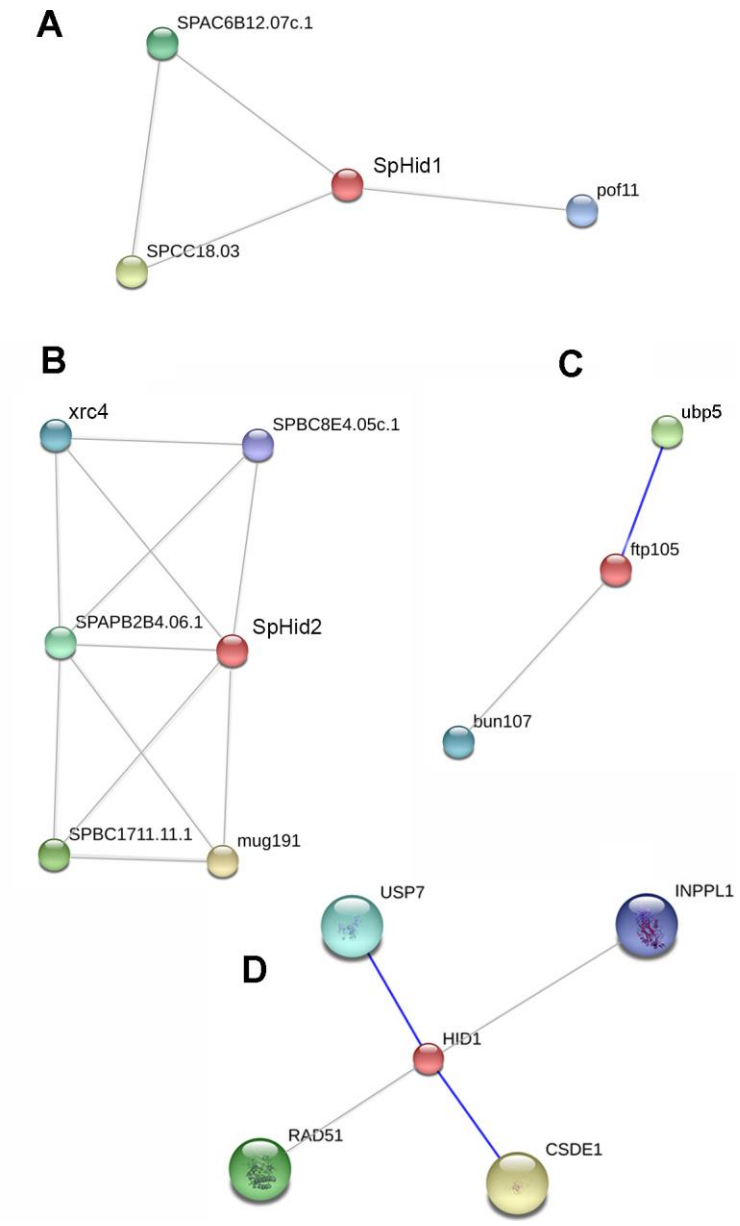


Figure 3.5. STRING output for S.pombe and human HID interactions. The interactions were pulled from the STRING database by search term using either the gene-specific identifier SPAP27G11.12 for SpHid1 (A) and SPBPG19A11.11c for SpHid2 (B) or the gene name in the case of S. pombe ftp105 (C) and the human HID1 (D). The coloured circles (nodes) represent specific proteins or genes and the connecting lines (edges) represent a potential interaction. Only interactions representing direct experimental evidence (dark blue edges) or gene co-expression (gray edges) related evidence are shown. The confidence cut-off used for the association was 0.4.

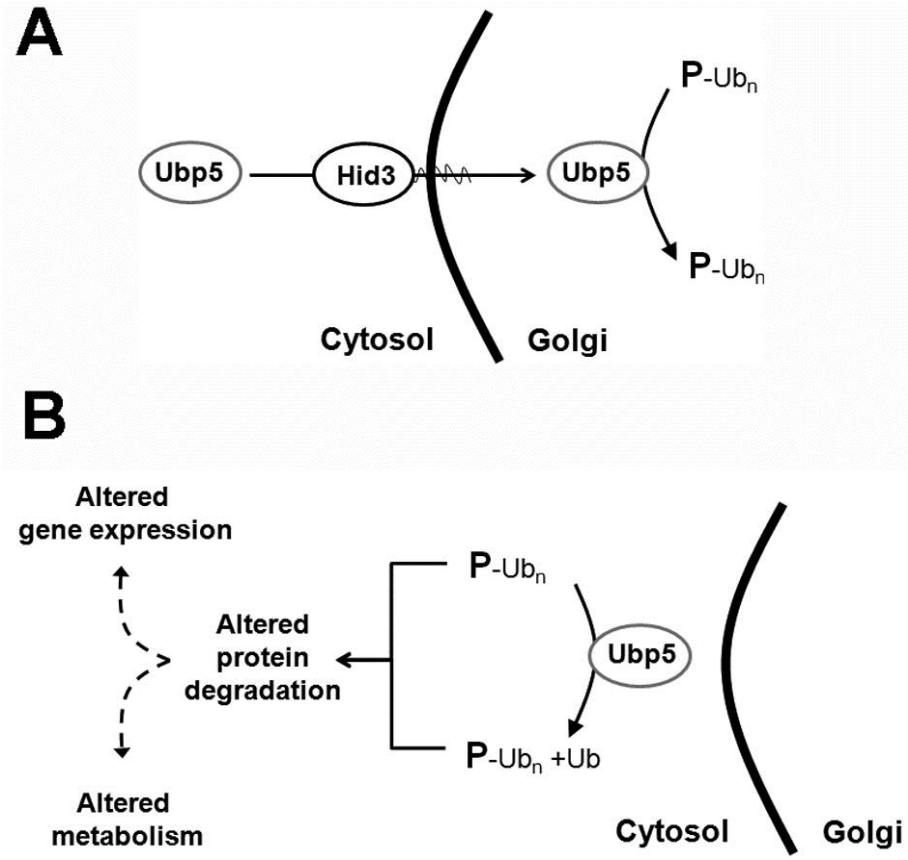


Figure 3.6. Model of Hid3 function. The known role of Hid3 is to facilitate the transport of Ubp5 into the Golgi apparatus (A). Upon elimination of Hid3, Ubp5 remains in the cytosol where it may remain active in the process of protein deubiquitination. This activity would expect to alter directly (solid arrows) protein levels and indirectly (dashed arrows) gene expression and metabolism.

3.2.2.3. Prediction of genetic and physical interacting partners

STRING database v8 (Jensen et al., 2009) was the starting point for investigating potential interacting partners for the 3 Hids proteins (Figure 3.5). This database has had frequent updates and so has been continually searched for relevant information (Szklarczyk et al., 2015). Other potential interactions, both physical and genetic have been explored through PomBase (Wood et al., 2012). STRING offered a convenient and logical way of searching for potential interactions by pulling together information from experiments to text-based associations. In order to develop working hypothesis for Hid proteins as strong as possible, associations based only on either co-expression or direct experimental evidence was considered. For Hid1 and Hid2, only interactions based on co-expression have been considered, since there is no current experimental evidence for any physical interaction. Gene *hid1* has been noted to show co-expression with two uncharacterized genes *SPAC6B12.07c* and *SPCC18.03*, which appear to have ubiquitin ligase activity and zinc ion binding activity (a PomBase notation). It also co-expressed with *pof11*, part of the ubiquitinatin ligase complex and for which there is evidence that the encoded protein binds Skp1 (Figure 3.5A). Gene *hid2* is co-expressed with mug191, a potential mannosidase involved in cell wall process, two uncharacterized genes, *SPAPB2B4.06* and *SPBC8E411.08c*, potentially associated with the mitochondrion, an uncharacterized gene *SPBC1711.11* potentially involved in vesicular transport and autophagy, and *xrc4*, which is involved in DNA repair, the protein of which has been localised to the nucleus and cytosol, like Hid2 (Figure 3.6B).

STRING showed the known physical interaction of Hid3 with Ubp5, which is a deubiquitinase that is located in the Golgi through the action of Hid3 (Figure 3.5C). The other gene was *bun107* that encodes a WD repeat protein that is involved in the positive regulation of ubiquitin protease activity and vesicle secretion. The human orthologue of Bun107 is WDR48, which is necessary for activation of deubiquitinating enzymes. The human HID1 was also queried (Figure 3.5D). USP7 is the orthologue of Ubp5 and so was shown to be associated based on the experimental evidence obtained from pombe. CSDE1 is a cold-shock domain-containing protein for which evidence suggests is an RNA-binding protein (Grosset et al., 2000). INPPL1 is a phosphoinositol phosphatase that negatively regulates the phosphoinositol signaling pathway (Pilot-Storck et al., 2010), but the evidence for this association is indirect. The association of HID1 with RAD51 is unlikely and has been pulled up through an overlap in nomenclature with DMC1 involved in DNA repair. A STRING search with the *C. elegans* hid-1 protein also pulls up interactions based on experimental evidence with a few proteins involved in

ubiquitination processes, such as the E3 ubiquitin ligase wwp-1 and the deubiquitinase math-33 (data not shown).

The disruption of proper localisation of Ubp5 in *ftp105Δ* indicates that Ftp105 is capable of physical interaction with it (Kouranti et al., 2010). It is not known if Ftp105 can physically interact with other proteins, like those mentioned above. The other type of interaction is genetic, where alteration in expression of the protein will affect the expression or activity of other proteins necessary for normal cell function. This genetic interaction can also be extended to the idea that the expression of two genes contributes to a phenotype and that such interaction can be investigated by determining the effects on growth of double mutant genotypes (Schuldiner et al., 2005). The basis of these types of genetic interaction studies is to identify potentially physically interacting partners (Kelley and Ideker, 2005). Two major studies have presented genetic interaction maps that have been summarized in Pombase (Roguev et al., 2008; Ryan et al., 2012). There are six interactions for *hid1* only one of which is a positive genetic interaction with SPAC1952.06c, which is a predicted member of the spliceosomal complex (Ryan et al., 2012). Ftp105 was included in the original epistasis map construction due to its potential involvement in chromosome structure (Roguev et al., 2008). Of the 101 genetic interactions identified only 27 were positive. The only conclusion that can be made is that because the vast majorities are negative, a mutation in *hid3* is more likely to have a detrimental than a positive effect on growth.

3.3 DISCUSSION

The purpose of this bioinformatic study was to determine at the start if the S. pombe HIDs represented good target to study as models for the function of HID1 in humans. This could also relate to understanding DYM function as well, because HID and DYM proteins show common structural features at the level of protein sequence. The first objective was to determine if HIDs had a similar evolutionary background to the DYMs that might indicate similar function. It is clear that both HIDs and DYMs came into existence with the appearance of eukaryotic organisms. The early appearance of eukaryotes was a time of a great genetic explosion when gene families specific to eukaryotes arose (Koonin, 2010). The phylogenetic comparison of HID and DYMs across the range of eukaryotic taxa revealed that the two proteins have quite different evolutionary paths. It is apparent that the lineage of neither protein conforms to the current idea of eukaryotic supergroups (Baldauf, 2003) and which corresponds to poor support for this idea in recent molecular evolution studies (Yoon et al., 2008). The major findings

from the phylogenetic study were 1) the confirmation that fungi have only HID and that fungi have multiple paralogues of HID, 2) that the Viridiplantae have only DYM, 3) some taxa have neither, and 4) evidence for horizontal gene transfer within the DYM superfamily.

Having DYM or HID is not a requirement for a cell being eukaryotic as shown by a vast number of organisms not having DYM or HID. Although, we have not identified a multicellular eukaryote with the exception of species of linked algae, that do not have one or the other. Thus, it is interesting to speculate that having either is necessary for cellular communication that makes multicellularism possible, but it also suggests there may be functional overlap. The major differences that have been noted between the lineages of the two proteins, such as the HID and DYM present within the choanoflagellates, the presence of DYM within the fungal family Mucorales, and the apparently different evolutionary rates of the HIDs and DYMs within fungi and plants, respectively, directly suggest that horizontal gene transfer took place at some point within their evolutionary histories, which is known to happen between eukaryotes (Keeling and Palmer, 2008). Species of Mucorales are parasitic and they can infect animals, plants or each other. For example, species of absidia can infect animals or plants and species of parasitica can infect and pass or retain genes from the absidia (Hoffmann et al., 2013; Kellner et al., 1993). Because the DYMs are restricted to the mucorales, it is interesting to speculate that they may have obtained them from another eukaryote, in this case from a non-fungal Opisthokont. Far reaching speculation is that they subsequently transferred them to the Viridiplantae, noting that the plant DYMs are more evolutionarily conserved to animals than the HIDs are within fungi. The phylogeny is an ongoing process that will require more detailed analysis of sequence properties. There are a number of sequences that fell below the criteria for selection within the global phylogenetic analysis that could provide more information as to early common features between the HIDs and DYMs. Also, it would be useful to try to root the trees using a more divergent sequence, such as c10orf76, which is the most similar protein pulled up by the BLAST searches.

The phylogenetic analysis confirmed S. pombe to have 3 HIDs. They most likely evolutionary model is that there was one HID ancestoral gene in fungi and that this diverged to produce the ancestor of *hid2* and the ancestor of *hid1* and *ftp105*, with the latter two having duplicated more recently. Using conventional nomenclature Ftp105 will now be called Hid3 in subsequent chapters. Although the predicted protein topology of Hid one appears more similar to that of the HID-1 of *C. elegans*, it appears that Hid3 retained the interaction properties that allow it to associate with the Golgi membrane (Kouranti et al., 2010). HIDs from both vertebrates

and invertebrates appear to associate with the Golgi membrane in a dynamic manner through N-terminal myristoylation (Mesa et al., 2011; Wang et al., 2011) so protein topology has little to do with membrane interaction. After the duplication of *hid1* and *hid3*, Hid1 was retained and its functional diversification to a cytosolic location provided some advantage to fungal growth. The known localisation of Hid3 is consistent with a dynamic shuttling between the cytosol and Golgi membrane with its interaction helped by anchoring through a fatty acid. N-myristoylation occurs at Gly2 once the terminal methionine is removed. All S. pombe Hids start with M-G so it remains unclear as to why Hid3 alone is a Golgi protein.

S.pombe represents a good model to study HID function in that it can be studied without the potential complications of DYM being present. However, in order for it to have implications for medical research, it would have to still have some of the same interactions that are observed for human HID1. The STRING database search was useful in highlighting some aspects of similar function. Although the associations derived from one organism are used to suggest interactions for another, the network interactions confirm that homologues of interacting partners exist and that function is conserved. There is evidence that the animal HID1 physically interacts with CSDE1, which is an RNA-binding protein involved in post-transcriptional regulation of gene expression (Mihailovich et al., 2010) and cell morphology and dynamics (Kobayashi et al., 2013). The closest S. pombe homologue of CSDE1 appears to be Prp22, an ATP-dependent RNA helicase. Importantly, the link between HID function and ubiquitin-related processes appears to be conserved between fungi and animals through the direct interaction of Hid3 with Ubp5. For Humans, USP7 is the functional homologue of S. pombe Ubp5, thus the association of HID1 with USP7 in human cells could be investigated. I have proposed a simple model by which altering levels of Hid3 or its complete elimination by mutation would have effects on protein levels and indirectly on gene expression and metabolism (Figure 3.6). This model shows that Hid function needs to be studied at the level of both gene and protein expression and for its influence on metabolism.

Chapter 4

Effect of stress treatments on the growth of *hidΔ* mutants

4.1 INTRODUCTION

The HID1 protein of *C. elegans* has found to be part of the switching mechanism located in the Golgi (Kouranti et al., 2010), HID1 targeting to the Golgi in *C. elegans* is sensitive to BFA (Wang et al., 2011), and the Golgi structure appeared to be disrupted in *S. pombe* lacking Hid3. In addition, the growth of the *hid3Δ* mutant of *S. pombe* is compromised and so it was essential to check if the poor growth phenotype was exacerbated by external stresses.

HID-1 of *C. elegans* has been associated with Dense Core Vesicles and can be transported to pre-synaptic regions of the plasma membrane in neurons. It is responsible for the correct localization of the of GTPase RAB-27 and has been implicated in neuropeptide transport (Mesa et al., 2011; Wang et al., 2011). HID-1 has also been associated with the *medial-* and *trans*-Golgi of *C. elegans* where it appears to be attached to the membrane through a myristoyl anchor (Mesa et al., 2011; Wang et al., 2011). Thus, the evidence obtained with the *C. elegans* HID-1 suggests that when present in the Golgi it is primarily involved in anterograde transport of peptides or proteins. With the recycling of components from the plasma membrane to the Golgi and from the Golgi to the ER, it is very difficult to determine if a Golgi protein is involved specifically in anterograde or retrograde transport (Sáenz et al., 2009).

4.2 RESULTS

An indication that a protein is involved in a the cellular mechanism to respond to and to cope with external stresses is that the gene encoding it changes expression by a stress treatment. This can either be up-regulation if the protein is needed for a new function or to increase a process or down-regulation if a particular process must be de-repressed. This is different from processes that may be down-regulated because they are non-essential or even detrimental to the tolerance of a cell to stress, although this down-regulation may be tightly controlled and important for cell survival. Up-regulated or activated PKases and TFs and down-regulated or inactivated PPases are examples of a system of opposite type of control (Shiozaki and Russell, 1995; Zhou et al., 2010). The study of global gene expression in response to stress conducted by (Chen et al., 2003) presented a good resource for a first look into the stress response gene expression of each of the *hids*. They studied 5 different types of stresses through separate treatments: oxidative (H_2O_2), heavy metal (Cadmium), abiotic (Heat);

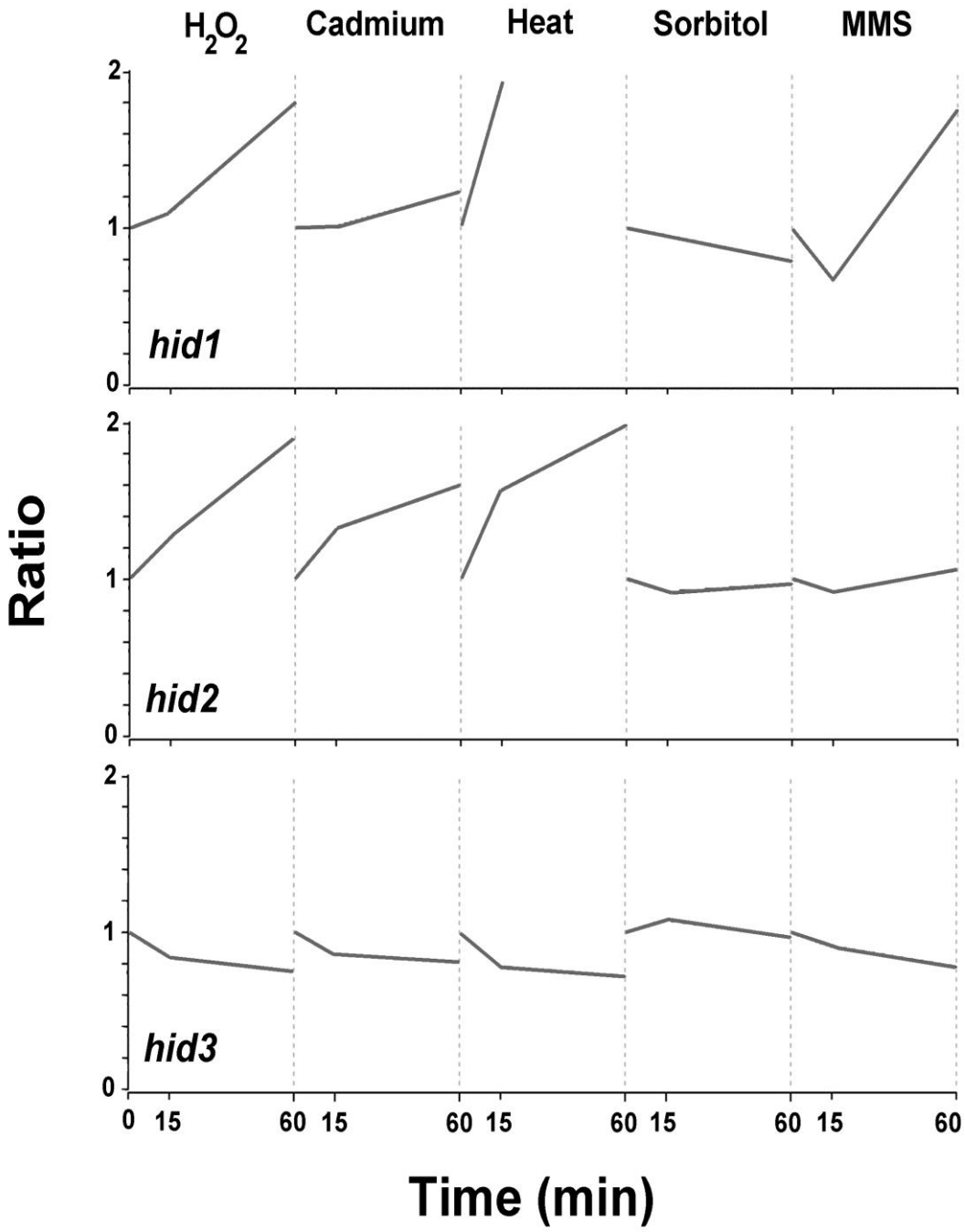


Figure 4.1. Relative expression of *hid* genes under stress. The values at times 15 and 60 min are relative to the zero time set to the arbitrary unit of 1. The figures were adapted from those taken from the Bähler laboratory website (http://128.40.79.33/projects/stress/) as described by (Chen et al., 2003).

Table 4.1. A summary of the effect of general stresses on the proliferation of mutant genotypes. The effect on growth was determined by cell count of cells incubated in YES liquid culture over period of 12h. Abbreviations are: TBZ, thiabendazole; BFA, brefeldin A; GA, golgicide A. The letters N and Y denote 'no' difference or an 'apparent' difference, respectively, compared to negative controls VCN and VCU expressing the marker gene. Each test was performed at least twice for each genotype, except where not determined (nd).

	Stress treatment						
Strain	NaCl	Heat	Cold	H_2O_2	TBZ	BFA	GA
*rad3*Δ	nd	N	nd	Y	nd	nd	nd
*bub1*Δ	N	nd	N	nd	Y	nd	nd
hid1Δ::natMX6 iso1	N	N	N	Y	N	N	N
hid1Δ::natMX6 iso3	N	N	N	Y	N	N	N
hid3Δ::natMX6 iso1	N	N	N	Y	N	N	Y
hid3Δ::natMX6 iso2	N	N	N	Y	N	N	Y
hid2Δ::ura4$^+$	nd	nd	nd	N	nd	nd	nd
hid2Δ::kanMX6 iso1	nd	nd	nd	N	nd	nd	nd
hid2Δ::kanMX6 iso2	nd	nd	nd	N	nd	nd	nd
hid2Δ::ura4$^+$ *hid1::natMX6 iso1*	nd	nd	nd	N	nd	nd	nd
hid2Δ::ura4$^+$ *hid1::natMX6 iso2*	nd	nd	nd	N	nd	nd	nd
hid3Δ::kanMX6 iso1 *hid2Δ::natMX6* iso1	nd	nd	nd	Y	nd	nd	nd
hid3Δ::kanMX6 iso1 *hid2Δ::natMX6 iso2*	nd	nd	nd	Y	nd	nd	nd

Table 4.2. A summary of the effect of DNA damaging agents on proliferation of mutant genotypes. The effect on growth was determined by cell count of cells incubated in YES liquid culture over period of 12 hours for the following treatments: mitomycin C (MTC), camptothecin (CPT), hydroxyurea (HU) and methyl methane sulfonate (MMS). UV light treatment was done on colonies on petri dishes containing standard YES/agar media and monitored for at least 3 days following exposure .The letters N and Y denote 'no' difference or an 'apparent' difference, respectively, compared to negative controls VCN or VCU. Each genotype was tested at least twice with each stress treatment.

	Stress treatment				
Strain	UV light	MTC	CPT	HU	MMS
*rad3*Δ	nd	nd	Y	Y	Y
hid1Δ::natMX6 iso1	N	N	N	N	N
hid1Δ::natMX6 iso3	N	N	N	N	N
hid3Δ::natMX6 iso1	N	N	N	N	N
hid3Δ::natMX6 iso2	N	N	N	N	N
hid2Δ::ura4$^+$	N	N	N	N	N

osmotic (sorbitol) and DNA damage (MMS). The data showing the expression of each of the three hids is shown in Figure 4.1. The gene hid1 was up-regulated within the first 15 min in response to oxidative, heavy metal and heat stress. It was not up-regulated in response to osmotic stress and there was a delay in its up-regulation in response to DNA damage. The gene hid2 was also quickly up-regulated in response to oxidative, heavy metal and heat stresses, but remained unchanged in response to osmotic stress and DNA damage. The expression of hid3 was unchanged, if not decreased, in response to any of the stresses. It should be noted that no hid showed a dramatic increase in expression, but even a 2-fold increase observed for hid1 and hid2 suggested that these proteins play an active role in response to stress.

4.2.1. Effects of stress treatments on growth of cells lacking Hids

A role in cellular stress response was examined by testing the growth mutant genotypes under different stress treatments. Pombase lists no stresses for which *hid1* and *hid2* mutants have been tested, however, the *hid3* mutant was reported to be sensitive to salt and heavy metal stress (Kennedy et al., 2008). Heat was also tested, because HID-1 of C. elegans had been identified in the process of mutant was identified the mutant of the yeast orthologue of the Hids, ECM30, is also sensitive to a number of different treatments and stresses, such as HU, but appears to be temperature insensitive (www.yeastgenome.org). Therefore, the mutants were tested using NaCl and a variety of other stresses to see if they behave differently to wild type and VC genotypes. In certain case control strains were used in order to check that the stress condition was working. For the action of the microtubule-destabilising compound thiabendazole (TBZ), the spindle assembly checkpoint kinase mutant *bub1Δ* was used (Bernard et al., 1998; Kadura et al., 2005). To check the action of the various DNA damaging agents, such as hydroxyurea (HU), Camptothecin (CPT) and Methyl Methanesulfonate (MMS), the check-point defective mutant strain *rad3Δ* (Bentley et al., 1996) was used due to its hypersensitivity to these agents.

The stress treatments were divided into two categories, general stresses (Table 4.1) and DNA damaging agents (Table 4.2). For those stress that I tested, each genotype was tried at least twice to check for any substantial difference from either WT or VC strains. The finding have been summarised in the tables, since the only stress for which there was any noticeable difference was for *hid1Δ* and *hid3Δ* under oxidative stress. At an H_2O_2 concentration of 0.7 mM, the generation times for both *hid1Δ::natMX6 iso1* and *hid3Δ::natMX6 iso1* were increased 3-4 fold from about 3.5 hours to

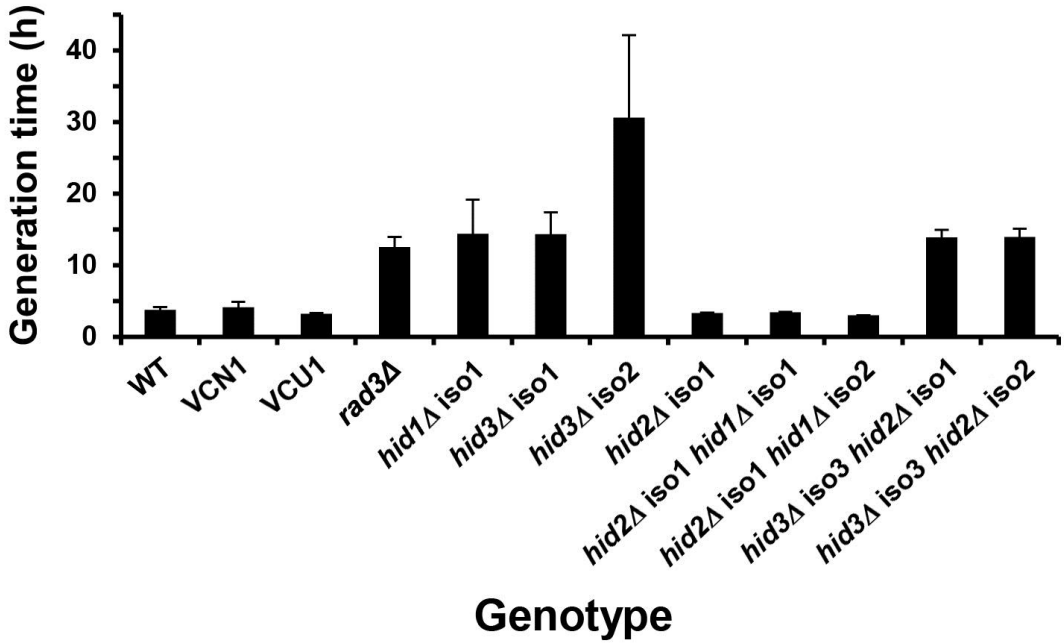


Figure 4.2 Effects of oxidative stress on mutant growth. Cells of each mutant strain were grown overnight in standard YES at 30 °C with shaking at 200 RPM. The cells were counted using the Cellometer™ Mini (Ozyme Ltd). A volume of the starter culture was added to 5 ml of YES medium containing 0.7 mM H_2O_2 to give a starting cell concentration of $5x10^5$ cells/ml. An aliquot was taken to count the initial culture and the rest of the culture was incubated for 12 hours under the conditions above. The generation time was calculated using the cell concentrations at 0 and 12 hours. Each strain was tested at least 3 times. Two-tailed, Student's t-tests for unequal variance were used to determine statistical significance between genotypes. The mutant *rad3Δ* was statistically different from WT at $P < 0.01$. The genotypes *hid1Δ::natMX6 iso1*, *hid3Δ::natMX6 iso1*, *hid3Δ::natMX6 iso2*, *hid3Δ::kanMX6 hid2Δ:: natMX6 iso1* and *hid3Δ::kanMX6 hid2Δ:: natMX6 iso2* were statistically different from both WT and VCN at $P < 0.01$. The other mutant genotypes were not significantly different from WT or the corresponding VC at $P = 0.05$.

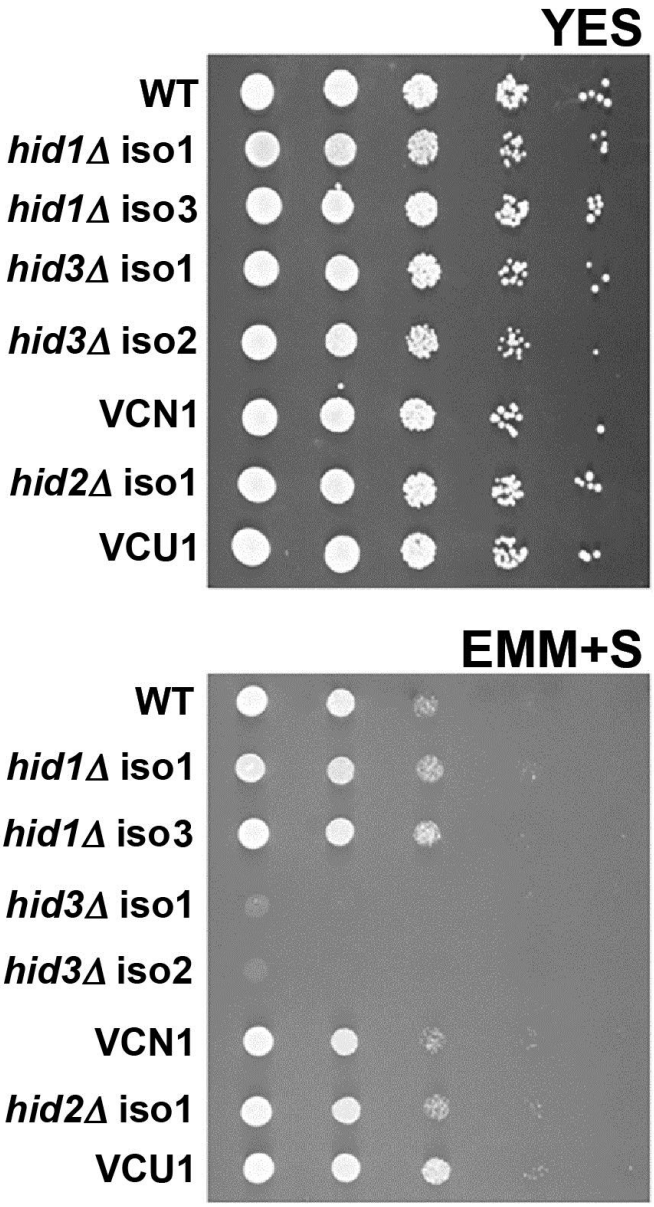
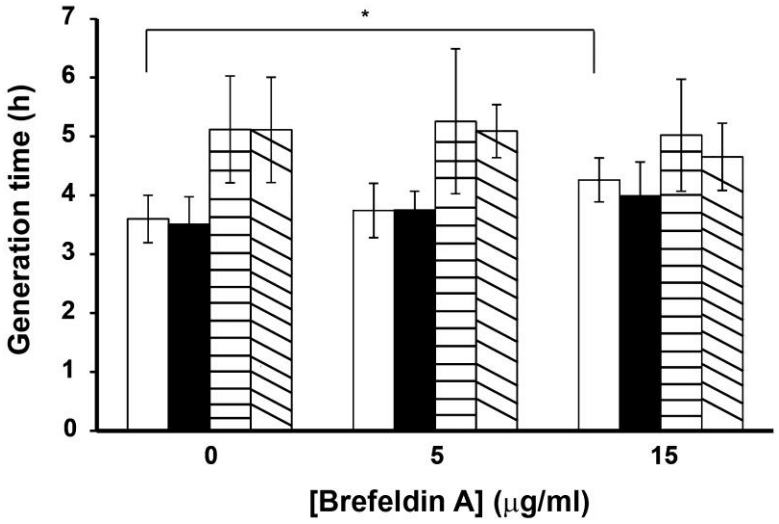


Figure 4.3. Differential proliferation of mutants on minimal medium. Cells of each mutant strain were grown overnight in standard YES at 30°C with shaking at 200 RPM. The cells were counted using the Cellometer™ Mini (Ozyme Ltd) and a series of 10-fold dilutions of cells were made starting from $5x10^6$ cells/ml (left column). A volume of 5 ml was placed for each spot onto petri dishes containing either YES/agar or EMM plus supplements. The EMM was purchased from Formedium Ltd (Norfolk, United Kingdom). Plates were incubated for 4 days before images were taken.

A



B

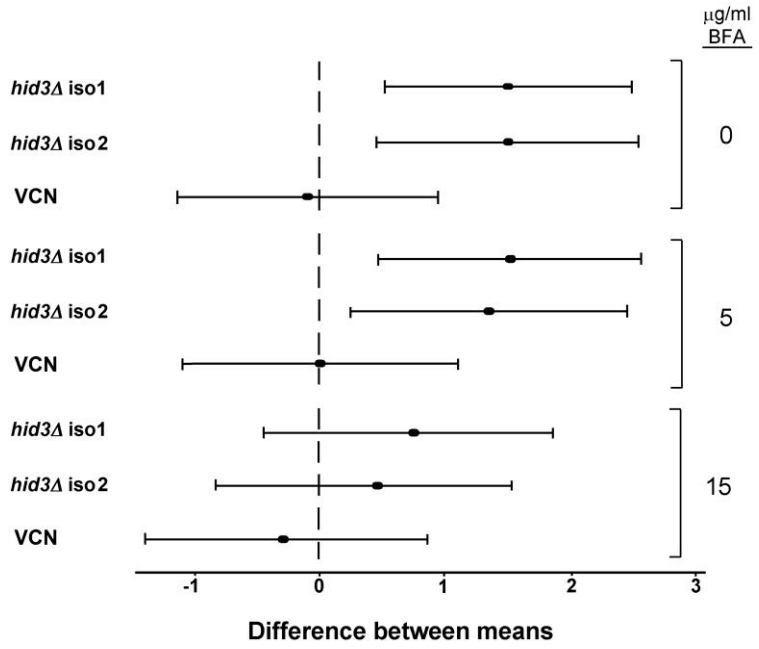


Figure 4.4. The effect brefeldin A (BFA) on mutant proliferation. (A) The dependence of generation time on BFA concentration. Bar representations: WT, open bars; VCN solid bars; *hid3Δ::natMX6 iso1*, horizontal lines; *hid3Δ::natMX6 iso2* sloping lines. Cells were prepared as described in the legend to Figure 4.2, but were incubated in the specified concentration of BFA for 12 hours. DMSO was added to 0 BFA cultures in the same volume used for adding BFA to treated cultures. Generation time was calculated from cell counts and statistical significance between samples was determined as described in the legend to Figure 4.2. Student's t-tests were conducted between the same mutant genotypes at the different concentrations. The only statistically significant different test was between WT at 15 vs 0 μg BFA ($P < 0.05$, $n = 4$ for all samples). (B) Dunnett's *post-hoc* test of significance between means. The statistical test was done using Minitab v 17. All sample means have been compared to WT at the respective concentration either 0, 5 or 15 μg/ml BFA. If the bar denoting a test between two samples crosses the line, it denotes a non-significant difference.

more than 12 hours (Figure 4.2). The *hid3Δ::natMX6 iso2* mutant proliferated slooly under these conditions that the calculated generation time could not be considered accurate as shown by the large variation in values. The *hid2Δ::ura4*$^{+}$ mutant was not sensitive to H_2O_2, and interestingly, when the *hid1* deletion was incorporated into the *hid2Δ* genetic background, the double mutant was not H_2O_2 sensitive like the *hid1Δ* single mutants. In contrast, the *hid3Δ hid2Δ* double mutants remained H_2O_2 sensitive.

Another stress that I had attempted to use was iron limitation. Initially, iron limitation was tested using the compound Ferrozine, which is a potent iron chelator, and when in the medium prevents iron from entering cells. Ferrozine was used at a concentration of 50 μM in both YES and EMM+S medium. There was no effect of Ferrozine on growth of mutant genotypes using YES, but it came unexpectedly that *hid3Δ* genotypes would not grow in liquid minimal-media cultures with or without Ferrozine. This growth defect displayed by *hid3Δ* in minimal media was compared with wild type (wt), Vector Control (VC) and other mutant genotypes using standard drop tests on agar plates containing only EMM with added supplements (Figure 4.3). On YES/agar media, all genotypes grew almost equally well with even the few cells plated at the lowest concentration producing colonies. In contrast, every genotype experiences some difficulty growing on minimal media alone and the *hid3Δ* mutants grew particularly poorly with cell growth only visible at the highest concentration of cells spotted.

4.2.2. *hid3Δ* mutants are resistant to Brefeldin A and sensitive to Golgicide A

Brefeldin A (BFA) and Golgicide A (GCA) are standard compounds used to disrupt Golgi function and are used, respectively, to discriminate between anterograde and retrograde protein transport through the Golgi. Since Hid3 is a Golgi membrane protein, the combination of the two compounds was used to determine if Hid3 worked preferentially in either direction. Preliminary tests with at various concentrations of BFA showed that concentrations of 25 μg/ml or more were lethal to cells and increased estimated generation times to 17 h or more for all genotypes. Therefore, a range of BFA concentrations from 0 μg/ml to 15 μg/ml were used in order to try to observe differences in growth among genotyopes. There was no significant effect of BFA on WT generation time at 5 μg/ml, but the average generation time increased slightly. There was a statistically significant increase in generation time for wild type at 15 μg/ml (Figure 4.4A). Although not significant, the VCN also showed this trend of increasing generation time with increasing concentration of BFA to 15 μg/ml. I observed the increase in generation time of *hid3Δ* mutants under standard growth conditions that had already been reported (Ashehri 2015).

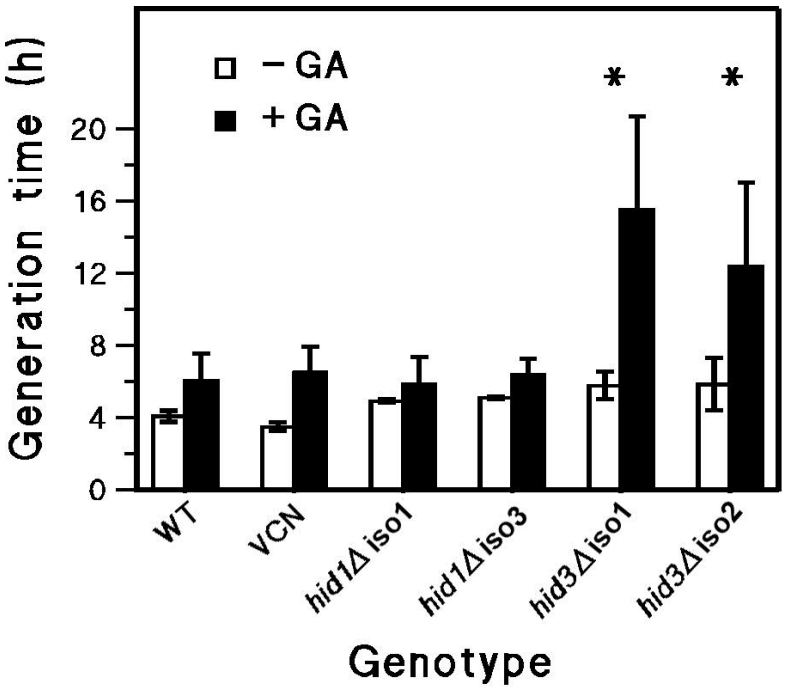


Figure 4.5. The effect of Golgicide A (GA) on mutant proliferation. (A) Comparison of generation times in the absence and presence of GA. Preparation of cell pre-cultures, treatment (but with 50 µg/ml GA in DMSO), and calculation of generation time were done as described in the legends to Figures 4.2 and 4.3. Statistical difference between means for each genotype with and without GA were determined using Student's t-tests. This was not done for VCN with n = 2. Only for the *hid1Δ::natMX6 iso1* genotypes was there no significant effect of GA (threshold P = 0.05, n = 3). The stars indicate genotypes with statistically significant different means with and without GA based on a Tukey's post-hoc test (n = 5 for WT and *hid3Δs*). (B) Dunnett's post-hoc test for significantly different means as described in the legend to Figure 4.4

In contrast to WT and VCN, growth of the *hid3Δ* mutants was not affected by the lower concentration of BFA.

In order to increase confidence that the differences in growth were, a different statistical test was used. The Dunnett's test compared differences in means in relation to one particular sample, i.e. WT at each of the concentrations of BFA (Figure 4.4B). Dunnett's test bases its statistical calculation on t-tests between sample means, but does it by taking into account a common variance among samples and provide confidence limits for non-overlap of means. The interpretation of a Dunnett's test is that two sample means are considered to be non-statistically different if the arm in any direction crosses the zero-difference point. Dunnett's test showed that VCN is very similar to wild type over all concentrations, but that the means of the generation times between each of the hid3*Δ* mutants and wild type become similar only at the concentration of 15 µg/ml. This showed that wild type and VCN generation times are increasing, but that hid3*Δ* generation times are, at least, changing less dramatically than those of WT and VCN. From Figure 4.4A, it appeared that generation times for *hid3Δ* were not changing with BFA. The *hid1Δ* mutants were tested and they were no different to WT and VCN.

Like for BFA, when used at too high of a concentration (< 100 µg/ml) GCA completely abolished growth of all genotypes. A concentration of 50 µg/ml was determined to give substantial differences between the *hid3Δ* mutants and other genotypes (Figure 4.5). At this concentration, the generation times for wild type, VCN and *hid1Δ* increased, but by less than 2-fold. In contrast, the generation times of the *hid3Δ* mutants increased on average by more than 2-fold. As mentioned previously, it was very difficult to get an accurate determination of generation time due to very little differences in cell numbers between 0 and 12 hours. Based on Student's t-tests, the effect of GCA was significant for wild type, VCN and *hid3Δ*. A Tukey's test of significantly different means, which is a conservative test taking into account a common variance among sample means, gave only the means of *hid3Δ* +/- GCA as being significantly different within genotype.

4.3 DISCUSSION

I initiated a series of test for responses to stress treatments based on the evidence that two genes orthologues to the HID1 genes of human and *C. elegans* appeared to be induced in response to various stresses (Chen et al., 2003). The variety of mutant and control genotypes constructed in the lab provided a good opportunity to determine if any of the Hids play a role in helping S. pombe cope with stress. Unfortunately, the delay in obtaining multiple alleles of the *hid2Δ* did not allow me to exhaustively test that mutant genotype or the double mutants lacking

Hid2. The treatments tested on *hid1Δ* and *hid3Δ* covered a wide variety of abiotic and mutagenic stresses. In general, the lack of Hid1 does not appear to be detrimental to cell growth in response to abiotic and mutagenic stresses, but it does seem to be necessary for normal growth in the presence of the high oxidative stress presented by incubation directly with hydrogen peroxide. Reactive oxygen species, like hydrogen peroxide, can act as a DNA mutagen as well as induce or reduce factors that may increase cell growth and proliferation that are hallmarks of cancer progression (Klaunig et al., 2010). The observation that *hid1Δ* is not sensitive to other DNA damaging agents suggests that Hid1 would most likely be involved in cellular homeostasis of general processes, such as metabolism or nutrient uptake. With very little data on the *hid2Δ* mutant, it is not possible to comment on why the sensitivity of *hid1Δ* to H_2O_2 treatment is reversed by removing Hid2.

The *hid3Δ* mutant is also insensitive to abiotic stresses and to DNA damaging treatments, but is sensitive to H_2O_2 treatment. The simplest explanation for *hid3Δ* sensitivity to oxidative stress is that the cells are already compromised in core metabolic or transport processes, because the Golgi apparatus is already disrupted. It may also be that stress signalling itself is dramatically altered if the Golgi itself is not functioning properly as suggested by ultrastructural studies of *hid3Δ* mutants (Alshehri 2015). Stress signalling components located in the plasma membrane are dependent on anterograde transport through the Golgi, but also recycle back to the Golgi, such as the GTPase Ras1 of S. pombe (Chang and Philips, 2006)

It was unexpected the *hid3Δ* mutants would not grow in EMM+S with glucose as the sole carbon source. It had been demonstrated that strains lacking this gene could be obtained using replacement with *ura4*$^+$, which requires selection in minimal media without uracil (Kouranti et al., 2010). However, colonies of *hid3Δ::ura4*$^+$ were quite small. The type of EMM used did make a difference on growth. All genotypes grew so poorly in EMM produced by the US company Sunrise Science Products Ltd that it wasn't possible to compare growth differences. Only the formulation produced by the UK company Foremedium gave the result shown in Figure 4.3. The effect wasn't due to potentially inhibitory properties of high glucose, since substitution with the other metabololisable sugars galactose, raffinose and fructose did not reverse the phenotype. The poor growth phenotype could be gradually alleviated by titrating in increasing amounts of YES, which would suggest that something within the EMM is missing to support *hid3Δ* growth. Although I have not investigated this greatly, it does not appear that amino acids or nitrogen are lacking as addition of particular amino acids, such as glutamine, also did not rescue the phenotype. It could be that *hid3Δ* is experiencing a quiescent state but with an extended

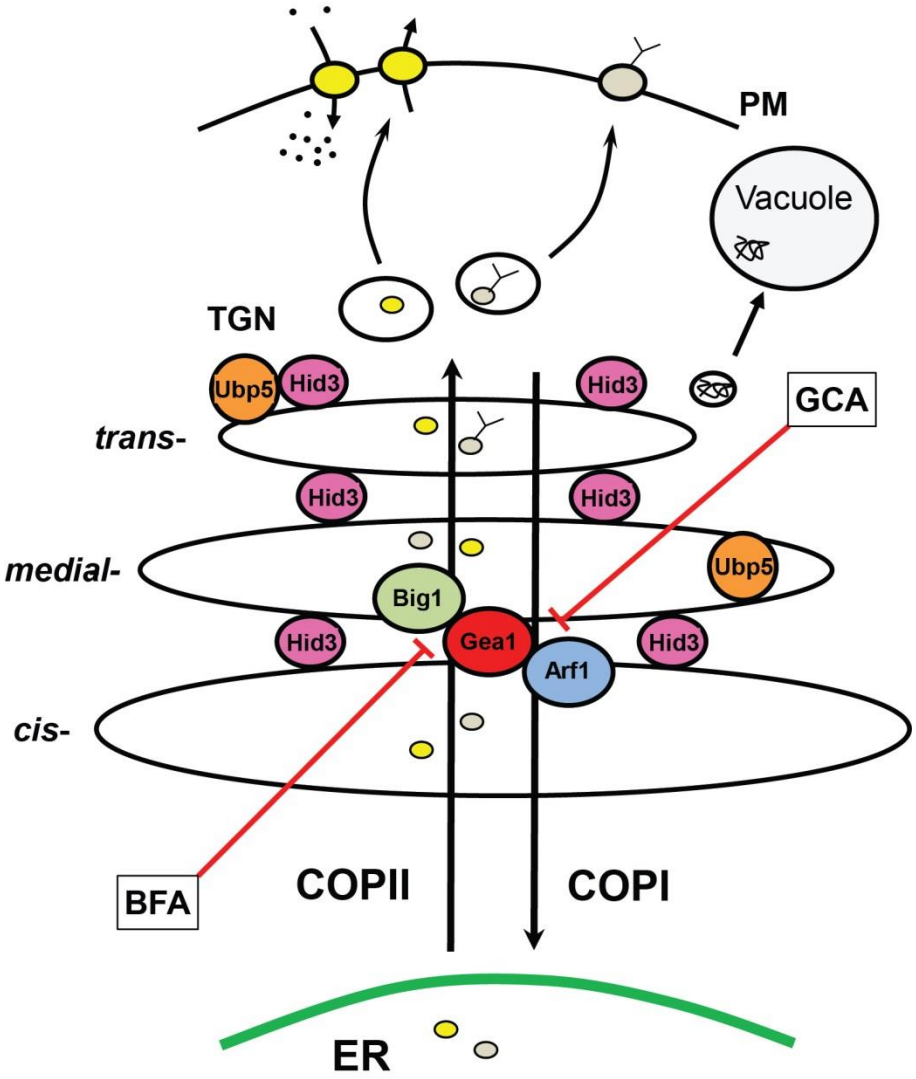


Figure 4.6. Hid3 involvement in protein transport through the Golgi Apparatus. COPII complex directs protein traffic from ER Golgi. COPI complex mediates retrograde transport from Golgi to ER. Proteins (coloured circles) pass through the hierarchal series of Golgi apparatus divisions where they are can be modified and transported to the plasma membrane or secreted. The forward progress of protein movement from ER through the Golgi for transport is called anterograde protein trafficking, which is mediated by the interaction of Gea1 with Big1. The return of protein components through the Golgi or from the Golgi to the ER is known as retrograde protein trafficking and is mediated by the interaction of Gea1 with Arf1. Anterograde and retrograde protein movement can be prevented by inhibition of Gea1 interaction with Big1 or Arf1 using BFA or GCA, respectively. I propose that Hid3 is responsible for helping to maintain the structure of the medial- and trans- Golgi in order to allow complete anterograde protein movement to occur. Ubp5 is also shown because it has been shown to be recruited to the Golgi by Hid3.

lifespan. It would be interesting to investigate the fate of individual cells in minimal media to see if they were dying or in a state of quiescence. One important result of the EMM effect is that a phenotype is now available in order to test if the CeHID and HsHID, and even the CeDYM and HsDYM, can complement the phenotype. If so, this would greatly support using S. pombe as a good model system for detailed mechanistic studies of these proteins as suggested in Chapter 3.

The complementary use of BFA and GCA has provided strong evidence that Hid3 works in the direction of anterograde protein transport. In mammalian cells, the action of both BFA and GCA is to inhibit GBF1, the guanine nucleotide exchange factor (GEF) that activates the ADP-ribosylation factor as part of the small GTPase signalling component of the Golgi apparatus. GBF1 is interesting in that it can either interact with BIG1/BIG2 to promote anterograde protein transport, or interact with the small GTPase Arf1 in order to promote retrograde transport (Bonifacino and Glick, 2004). BFA and GCA inhibit the action of GBF1, but in the anterograde and retrograde directions, respectively (Sáenz et al., 2009).

The observation that *hid3Δ* is unaffected in response to BFA at low concentrations whereas WT is not suggests that the mutation mimics the effect of BFA. The opposite effect of GCA to stop *hid3Δ* makes sense in that GCA combined with the lack of Hid3 would completely disrupt protein trafficking through the Golgi in both directions. Figure 4.6 presents a conceptual model of how Hid3 placement within the medial- and trans-Golgi network acts as an anchor to stablise Golgi structure to allow the movement of proteins through the Golgi network. With *S pombe* possessing orthologues of all the major factors, it is a very good model for studying Golgi protein trafficking (Eckler et al., 2013). It is interesting to speculate that Hid3 may be responsible for recruiting one or more of these factors to the Golgi as it does for Ubp5.

Chapter 5

Metabolomic and proteomic characterization of the *hid1Δ* and *hid3Δ* mutants of *S. pombe*

5.1 INTRODUCTION

Proteins are the major mediators of the cellular phenotype and function. Therefore, there has been intense interest to understand how protein levels change in response to cell perturbations. Unlike genetic mutation, which affects one gene, the consequences of stress or genetic mutation can result in changes to hundreds of the proteins in a cell. This is a hallmark of cancer, whereby a one or a few genetic mutations that start cell proliferation can have wide-spread effects on protein levels and ultimately cell metabolism (Dang, 2012; Jones and Thompson, 2009). Likewise, a cell's complement of proteins will change due to environmental factors, drugs, stress, growth conditions and extracellular signals. Like for gene expression, efforts continue to focus on how proteins in a tissue or organism change on a global level, where the function of a protein can be evaluated in the context of the other proteins that are present.

The term used to describe studies where as many proteins in a cell, tissue or organism are detected and quantified is called proteomics (Washburn, 2011). Similar to proteomics, metabolomics is the study of the cellular or tissue metabolites. It is a powerful tool to judge the metabolic state of a cell, such as its energy demands and nutrient requirements (Griffin, 2004). In general, gene expression is the easiest to do and the entire predicted transcriptome of a cell can be tested using microarrays or RNAseq (Perdacher, 2011). Proteomics suffers the lack of the broad detection of proteins likely to be present in a cell. However, extraction techniques using organic solvents like phenol or ether can do a good job to extract most cellular proteins, both soluble and membrane bound. Proteins are detected and quantified using 2-D gel electrophoresis or one of the quantitative mass spectrometry techniques, like ITRAQ, SILAC or label-free (Washburn, 2011). Label-free is becoming increasingly popular for proteomics. It has advantages over techniques involving labeling of proteins or peptides for mass spec detection, such as ITRAQ and SILAC/SILAN. It is less complex experimentally, less costly, and it has no limits to the number of samples compared like 2-D gel electrophoresis (Lassout et al., 2014; Strathmann and Hoofnagle, 2011).

Getting an overall picture of metabolism through metabolomics is probably the most difficult, because metabolites have many different chemical and physical properties so that one extraction technique or detection and measurement technique is insufficient to observe all classes of metabolites. In order to judge the energetic and nutritional state of the cell, usually the alcohol soluble metabolites, such as sugars, amino acids, organic acids and nucleotides. These can be observed and quantified using either proton nuclear magnetic resonance (NMR) or mass spectrometry (Kell, 2004).

Where there have been countless proteomic and metabolomic studies on human tissues, budding yeast and plants, relatively few have been done on pombe. This is particular so for metabolomics. Nevertheless, there have been both proteomic and metabolomic studies related to my project. Changes to the proteome of *S. pombe* have been investigated in normal and quiescent cells (Marguerat et al., 2012) with attempts to measure absolute levels of transcripts and proteins. They determined that protein molecules, even for signalling proteins, can exist in hundreds to thousands of copies with transcript levels being in the ones to tens. A combination of transcriptomic and proteomics was used to investigate gene expression regulation by non-coding RNAs (Leong et al., 2014). A parallel study using proteomics and proton NMR metabolomics was use to investigate metabolome and proteome changes in cells with altered stress signalling pathways under oxidative stress (Weeks et al., 2006).

I undertook a combination metabolomic and proteomic study in order to obtain a better understanding of molecular changes resulting from the elimination of the Hid1 and Hid3 proteins through gene deletion. Proteomics was considered to be appropriate from the evidence suggesting that the deubiquitinase Ubp5 is mislocalised to the cytosol from the Golgi in the *hid3* mutant (Kouranti et al., 2010). As stated in Chapter 1, relocalisation of a deubiquitinase from one compartment to another would likely have major effects on protein turnover. For the metabolomics we used proton NMR to quantify detectable ethanol soluble metabolites and a label-free LC-MS/MS approach to quantify proteins. The genotypes compared were WT, VCN, *hid1Δ* and *hid3Δ*. The metabolomics and proteomics revealed that *hid3Δ* cells appear to be stressed although they were grown under standard conditions. In *hid1Δ*, metabolic proteins were altered, mainly involving membrane-related processes. This suggests that Hid1 may be involved in maintaining metabolic homeostasis.

Table 5.1. Scheme of preparation of samples for post-genomic studies. Cell cultures were growth to a density of $5x10^7$ cell/ml. The cells were collected by centrifugation, immediately frozen in liquid N_2, and ground to fine powder using a cryomill. The genotypes were combined at this step, where equal masses of powder were combined for the two genotypes specified.

Sample	Combined Genotypes
WTa	WT
WTb	WT
WTc	WT
hid1Δa	*hid1Δ::natMX6* iso1 + *hid1Δ::natMX6* iso3
hid1Δb	*hid1Δ::natMX6* iso1 + *hid1Δ::natMX6* iso3
hid1Δc	*hid1Δ::natMX6* iso1 + *hid1Δ::natMX6* iso3
hid3Δa	*hid3Δ::natMX6* iso1 + *hid3Δ::natMX6* iso3
hid3Δb	*hid3Δ::natMX6* iso1 + *hid3Δ::natMX6* iso3
hid3Δc	*hid3Δ::natMX6* iso1 + *hid3Δ::natMX6* iso3
VCNa	VCN1 + VCN2
VCNb	VCN1 + VCN2
VCNc	VCN1 + VCN2

5.2 RESULTS

Table 5.2. Metabolite absolute values as determined by ^{1}H-NMR. Metabolites were extracted from approximately 30 mg of dried powdered cells. Metabolites were identified and quantified according to Moing et al (2004).

Metabolite	WT	*hid1*Δ	*hid3*Δ	VCN
Glycerol	122.4 ± 9.3	79.2 ± 4.5[b]	140.6 ± 9.8[b]	109.2 ± 2.4
Betaine	93.3 ± 1.2	101.0 ± 5.0	67.4 ± 0.6[b]	96.4 ± 0.4
Lysine	45.5 ± 0.7	63.8 ± 9.6	54.8 ± 4.5	57.8 ± 5.0
Adenosine_like	36.1 ± 0.8	37.5 ± 2.0	30.8 ± 0.1	36.3 ± 4.6
UnkS2.12	32.1 ± 2.4	33.6 ± 1.6	26.9 ± 0.6[b]	33.4 ± 2.8
Glucose	30.3 ± 5.8	6.3 ± 2.1	16.8 ± 7.9	8.5 ± 2.9
Glutamate	26.9 ± 1.3	35.0 ± 1.5	21.2 ± 0.8[b]	33.8 ± 1.4
Acetate	25.3 ± 2.9	18.0 ± 0.6[b]	24.7 ± 1.4[b]	21.0 ± 0.7
Alanine	24.5 ± 1.9	29.2 ± 2.0	30.7 ± 0.0[b]	28.6 ± 0.2
Leucine	15.2 ± 0.9	14.9 ± 0.4	12.2 ± 1.2	14.1 ± 0.9
Glutamine	14.1 ± 3.5	11.9 ± 0.6	12.0 ± 1.4	11.2 ± 0.4
Ac-choline[a]	12.5 ± 0.3	10.6 ± 0.2[b]	15.6 ± 0.9	13.0 ± 1.3
UDP_like	11.1 ± 0.6	13.3 ± 1.4	9.6 ± 0.4	10.8 ± 0.8
Succinate	10.8 ± 0.4	7.9 ± 0.1[b]	9.0 ± 0.3	10.0 ± 0.8
Trehalose	8.8 ± 0.8	6.2 ± 0.4	10.5 ± 0.8[b]	7.4 ± 0.7
UnkM5.80	8.2 ± 0.7	7.1 ± 0.3[b]	6.9 ± 0.7	6.4 ± 0.2
Aspartate	7.0 ± 0.8	6.8 ± 0.2	9.7 ± 2.1	7.6 ± 2.3
Valine	4.7 ± 0.2	5.7 ± 0.2	5.4 ± 0.2	5.7 ± 0.0
Formate	4.4 ± 0.7	3.2 ± 0.4[b]	2.1 ± 0.1	2.4 ± 0.2
Choline	4.0 ± 0.1	3.9 ± 0.1[b]	4.6 ± 0.1	4.8 ± 0.1
Nicotinamide	3.9 ± 0.3	4.1 ± 0.2[b]	4.5 ± 0.2	4.5 ± 0.1
Citrate	3.9 ± 0.1	4.1 ± 0.3	7.3 ± 0.1[b]	5.3 ± 1.0
Isoleucine	3.8 ± 0.5	3.3 ± 0.1	3.7 ± 0.5	3.6 ± 0.3
UnkM5.56	1.7 ± 0.1	1.4 ± 0.1[b]	2.9 ± 0.1[b]	1.7 ± 0.0
Phenylalanine	1.6 ± 0.2	1.9 ± 0.1[b]	1.3 ± 0.1[b]	1.7 ± 0.1
Tyrosine	1.3 ± 0.2	1.5 ± 0.2	1.4 ± 0.2	1.4 ± 0.1

a putative assignment

b significantly different ($P < 0.05$) from VCN based on an unpaired, two-tailed, Student's t-test assuming equal variance.

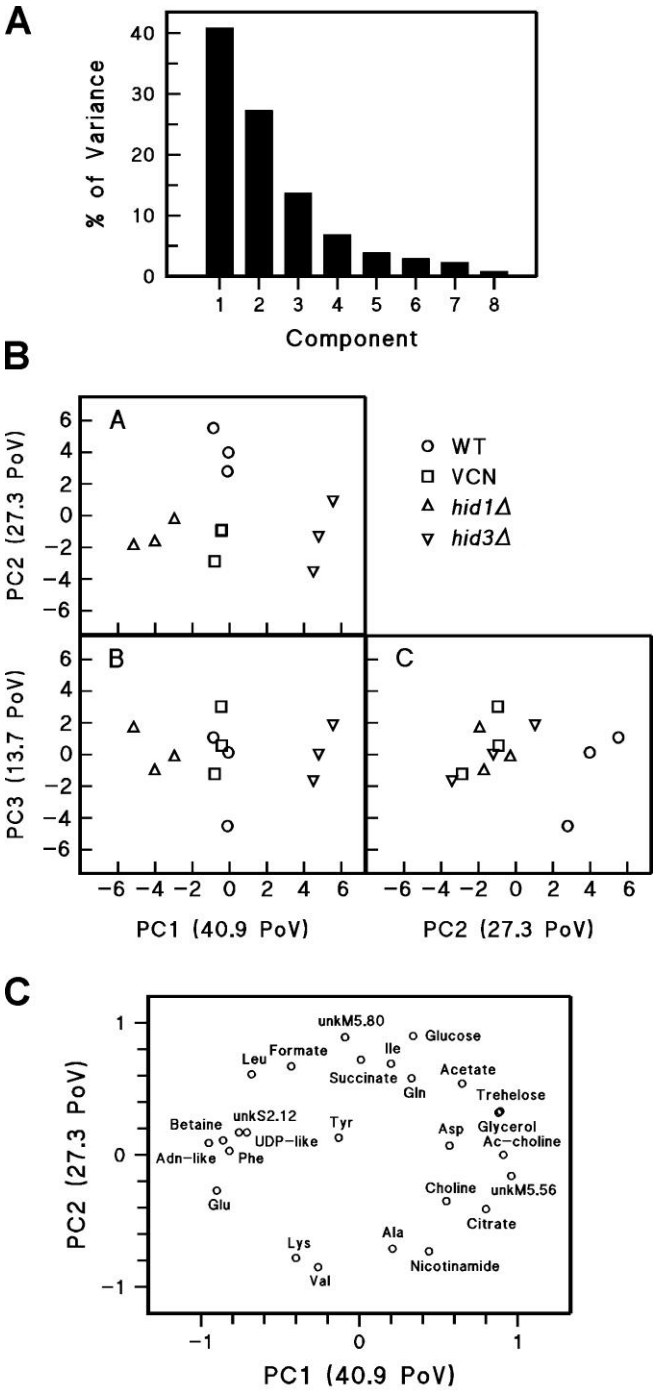


Figure 5.1. Global comparison of metabolite data among mutant and WT genotypes. Metabolite data was obtained by ^{1}H-NMR profiling. The same samples were analysed as used for the proteomic study. (A) Variance in data as a function of Principal Component. (B) Pair-wise scores plots using the first 3 PCs. PoV represents the Percentage of Variance. (C) Loadings plots describing the scores plot of PC2 vs PC1. The amino acids are represented by the appropriate 3-letter code. UDP stands for uridine diphosphate, Ac for acetate and Adn for adenine dinucleotide. Unknown compounds are represented by the number of peaks comprising the signal (S for singlet, M for multiplet) and the shift in PPM from the external standard.

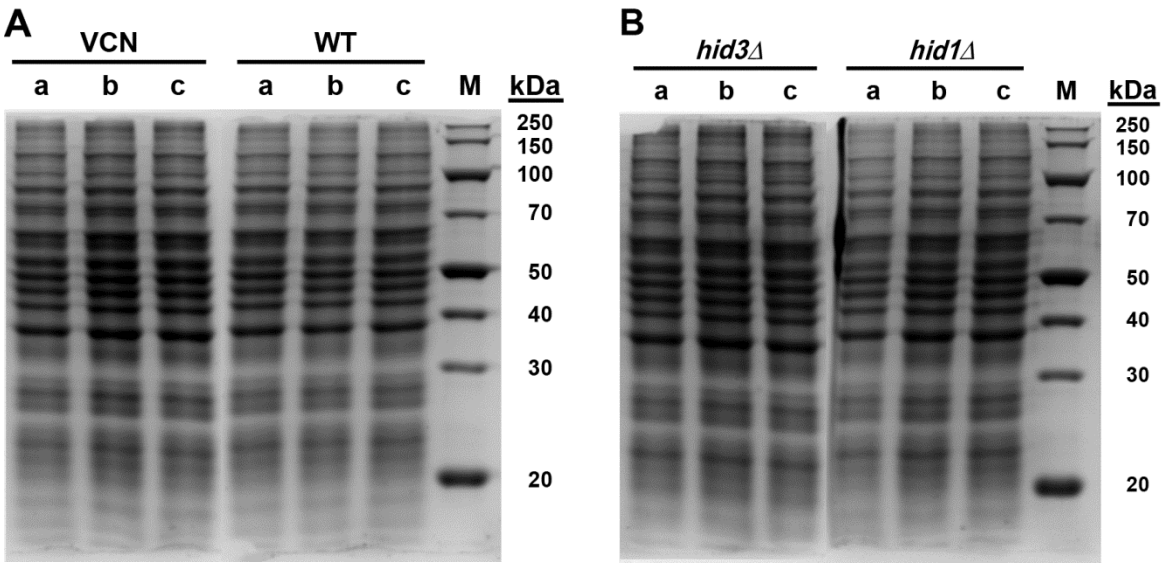


Figure 5.2. Denaturing PAGE of protein extracts used for proteome analyses. Gels of cell protein extracts from VCN and WT (A) and mutant (B) genotypes, respectively. Total protein in each extract was determine by the Bradford method using BSA as standard. A total of 25 μg of protein for each sample was prepared in Laemmli loading buffer and loaded onto a 10% SDS denaturing gel. After migration, the proteins were stained in QC Colloidal Blue (Bio-rad 161-0803. The VCN samples in (A) and the *hid3*Δ samples in (B) have been moved from the right side of the gel to the left side of the gel for the purposes of labeling the image. The molecular masses of the markers M are given on the right. The small letters represent an individual sample as explained in Table 5.1.

5.2.1. Strategy for sample preparation for metabolite and protein profiling

Cells were grown in 50 ml cultures starting with an initial density of $5x10^5$ cells/ml. Cells were grown for 16 to 18 hours until concentrations reached $5x10^7$ cells/ml at which point they were harvested. This cell density was chosen as it would likely reveal the biochemical differences among strains (Alshehri 2015; see Chapter 4). Cell growth in 50 ml volumes ensured that there was sufficient material to conduct all aspects of the post-genomic study. Potential biological variation within mutant and control strains was accounted for combining two independent mutation events for each genotype (Table 5.1). This scheme gave 12 samples from which polar metabolites and total proteins were extracted for ^{1}H-NMR and proteomic profiling, respectively.

5.2.2. Metabolic consequences of removing Hid1 and Hid3

The polar metabolites were extracted from dry powdered cell material using a deuterated aqueous buffer according to (A et al., 2004). It must be noted that the cells were not washed prior in order to freezing in order to minimise changes to the proteome and transcriptome, which were the priorities of the study. However, carry-over of metabolites from the YES media would probably serve to mask differences among the various genotypes, thus any identified differences would be real. The output from the analysis was 26 metabolites quantified absolutely on a μmol/gm dry weight basis (Table 5.2). It is apparent that the NMR could detect the added supplements, such as adenine (adenosine), uracil, (UDP-like), lysine, leucine and histidine (likely peak at S2.12) and glucose. These metabolites were high in quantity and they varied little among genotypes. The two most abundant metabolites were glycerol and betaine, which are two osmoprotectants and these were modified in each of the mutants compared to VCN. Each mutant genotype had about 10 metabolites that were significantly different to VCN, but differences were not very great.

Since there was little difference in individual metabolites, Principal component analysis (PCA) was conducted according to sample in order to see if any general metabolic differences among genotypes were visible (Figure 5.1). The plot of percent variance versus principal component showed that 82% of variance was represented the first 3 PCs (Figure 5.1A). Therefore, these three PCs were used for pairwise plots to reveal potential differences among genotypes (Figure 5.1B). The plot of PC3 vs PC2 shows the clustering of *hid1Δ*, *hid3Δ* and VCN showing that the first 3 PCs were sufficient to describe the variation between mutant and VCN samples. In the plot of PC2 vs PC1 there was no difference between wild type and VCN

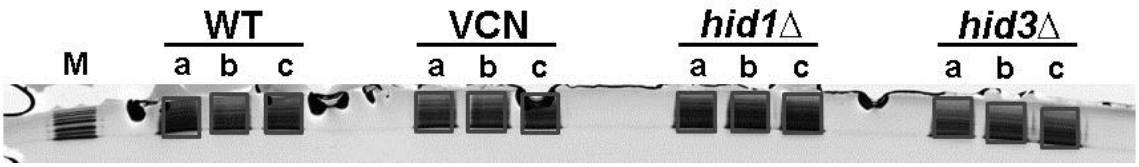


Figure 5.3. Denaturing PAGE preparation of samples for trypsin digest. Ten μg of total protein were loaded on to a 8% SDS-PAGE gel and migrated until all the proteins had entered into the gel. The square boxes (~0.5 x ~0.5 cm) represent size of the gel that was cut out for doing the in-gel trypsin digest. M stands for the molecular mass markers. The small letters represent individual samples for each genotype as shown in Table 5.1.

samples across PC1, which represented > 40 PoV. Both mutant genotypes were separated from VCN across PC1 showing that the metabolite profiles of the mutants were different to the controls, and they were split in opposite directions. All three modified genotypes were separated from wild type across PC2 and in the same direction suggesting that the similar metabolic changes were responsible. This difference in modified genotypes compared to wild type was not apparent in the PC3 vs PC1 with VCN clustering with WT. Therefore, the variance in PC2 was something common among the modified genotypes compared to WT with the most likely explanation being that expression of the marker gene did affect their metabolism. The loadings plot of the metabolites in the dimensions PC2 vs PC1 gave some indication of the metabolites responsible for the separation (Figure 5.1C). The metabolites separating *hid1Δ* were the added compounds adenine and uracil, the aromatic amino acids phenylalanine and the amino acids glutamate and betaine. Of these, none were significantly changed from VCN indicating that they likely served to distinguish the VCN from wild type. The separation of *hid3Δ* was characterised by the organic acids, citrate and acetate, the glycerol and trehalose and the amino acids choline and its acetate-modified form Ac-choline. Of these metabolites, only trehalose and citrate were significantly different to VCN indicating that the others were also defining the difference between VCN and WT. In addition, the VCN genotypes were characterized by the hydrophobic amino acids alanine and valine and the basic amino acid lysine contributed to the separation of VCN from wild type. Wild type was characterised by the other hydrophobic amino acids leucine and isoleucine, the organic acids succinate and formate and the sugar glucose.

5.2.3 Proteomic study on the effects of loss of Hid proteins

Total proteins for the proteomic analysis were extracted using an aqueous buffer with 1% triton detergent. It was expected that this protocol would give mainly soluble proteins and those proteins peripherally associated with membranes. After quantification of total protein, a pre-visualisation of the protein extracts was done by 1-D SDS-PAGE, which did not reveal any major difference in banding pattern among the genotypes (Figure 5.2). The gels also indicated that sample quality and total protein quantity were not greatly different among the samples, so that the proteomic analysis would be done on samples of similar properties. This would make down-stream preparation of samples and the data analysis more straight-forward. For example, it is easier for automated software, such as Progenesis IQ, to align spectra and normalise intensity values for various spectra with peaks of similar intensity. A total 10 µg of protein was used for label-free proteomic analysis with prior gel purification and in-gel trypsin digest (Figure 5.3).

Table 5.3. Summary statistics for proteomic analyses. The output from the Q-Exactive™ mass spectrometer was run through Proteome Discoverer™ vs 1.4 (Thermo Scientific) and peptide masses queried using the SEQUEST search algorithm against the P(roteome) S(et)_2014_01 for Schizosaccharomyces pombe, which hold entries for 5091 proteins.

Sample	No. Unique peptides	Number of proteins ID	number of proteins >1 unique peptide
WTa	11003	1883	1294
WTb	10425	1797	1262
WTc	10248	1791	1226
VCNa	10782	1815	1309
VCNb	10811	1841	1332
VCNc	10821	1872	1322
*hid1*Δa	10743	1818	1277
*hid1*Δb	10870	1855	1286
*hid1*Δc	10625	1808	1267
*hid3*Δa	10986	1843	1327
*hid3*Δb	11177	1878	1354
*hid3*Δc	10921	1836	1324
Mean	10783	1836	1296

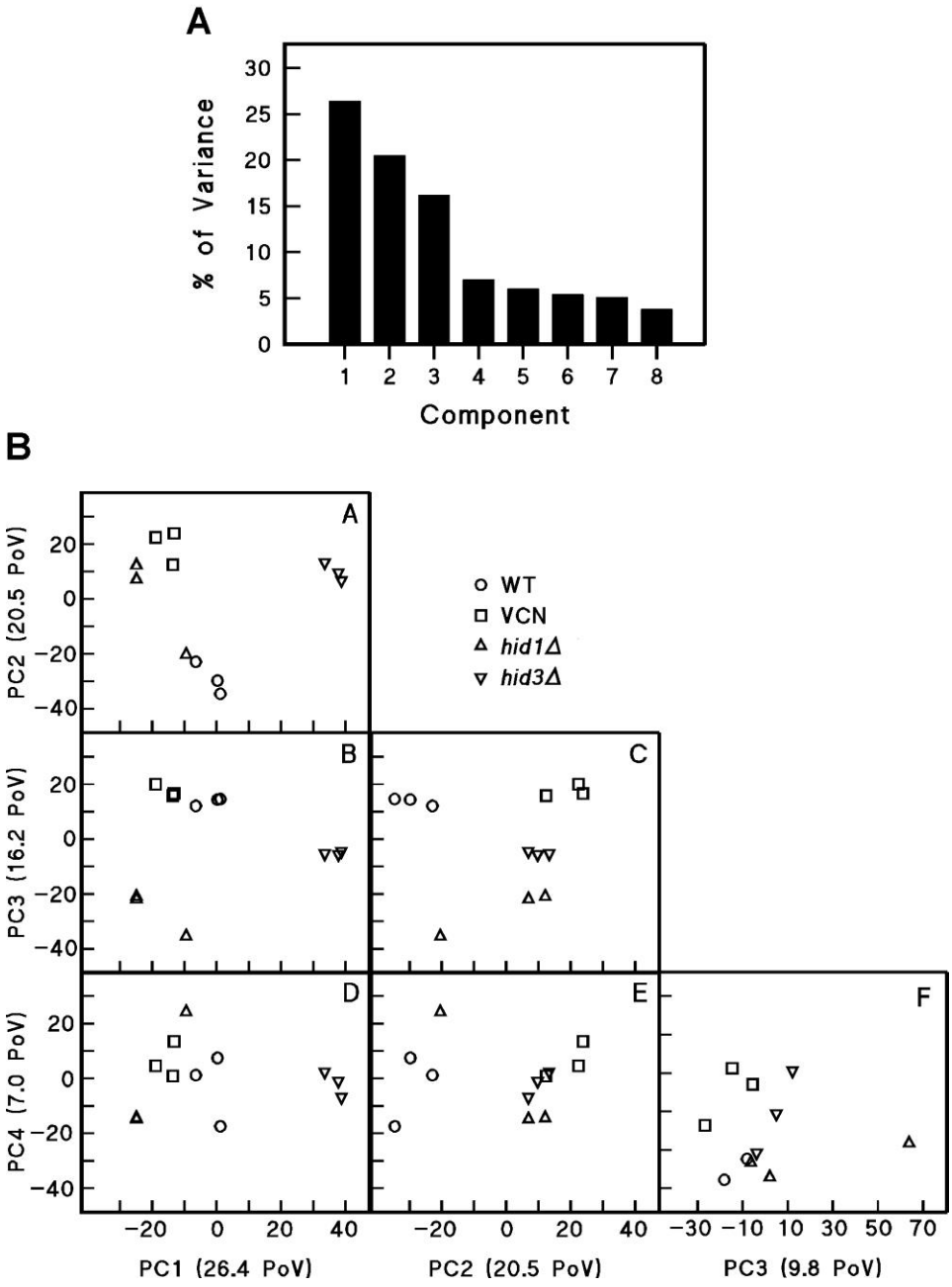


Figure 5.4. PCA of normalised values of quantified proteins. PCA was conducted on normalized intensities of common proteins as determined by the Progenesis software. (A) Variance in data as a function of Principal Component. (B) Pair-wise scores plots using the first 3 PCs. PoV represents the Percentage of Variance. The number of proteins used for the PCA analysis was 1911 after removal of proteins for which samples had missing data. The PCA analysis was conducted in R using the package FactoMineR.

The peptide mixture from the trypic digest was injected directly into a Q-Exactive™ mass spectrometer (Thermo Scientific) and proteins identified from peptides using Proteome Discoverer™ v 1.4 software. The information output from the mass spectrometer, in terms of the number of peptides quantified (average = 10783), the number of different proteins detected (average = 1836) and the number of proteins characterized by more than one unique peptide (average = 1296), indicated that the protein preparation and treatment of protein samples were quite uniform among samples (Table 5.3). On average, 1836 proteins were identified per sample, which represent approximately 36% of the predicted proteome, not including dubious and transposon sequences (www.pombase.org). This number of proteins was similar to the 1896 proteins detected in a study of nutrient stress responses to the *S. cerevisiae* proteome (Gutteridge et al., 2010) and the 2147 proteins detected in a study of the changing proteome in response to oxidative stress in S. pombe (Lackner et al., 2012).

Only those proteins identified by two or more unique peptides were selected by the Proteome Discoverer software for subsequent quantification using Progenesis LC-MS v 4.0. Proteins were quantified as the sum of the peptide intensities in order to give greater weight to most abundant peptides. The peak intensities output for each sample was normalised by a scalar factor to allow comparison directly between samples. The result was a list of 2058 proteins with normalised intensity values that was used for PCA of genotype differences (Figure 5.4). The variance in the protein data was spread much more evenly over the PCs with PC1 only covering 26 % of the variance. The first four PCs covered just over 70 % of the variance describing the data set (Figure 5.4A). However, it was clear that 4 PCs were sufficient to show differences among the genotypes as all samples clustered around the zero axis of the 4th PC (Figure 5.4B). In the plot of PC2 vs PC1, the *hid3Δ* samples were removed from the other genotypes across PC1 and *hid1Δ* clustered with VCN. This revealed that the the *hid3Δ* mutation had a greater effect on altering the proteome than the *hid1Δ* mutation. Similar to the metabolite PCA, the protein PCA also showed a separation in the second dimension where all modified genotypes were separated from WT. Again, I conclude that the presence of the marker protein has affected the overall proteome sufficiently for it to become apparent in the PCA.

Using the normalised values, Progenesis calculated the statistically relevant differences in each protein between the WT, mutants and VCN. The number of proteins UR and DR by more than 2-fold are shown in Figure 5.5. Compared to VCN, *hid1Δ* had 30 and 42 proteins UR and DR, respectively, whereas *hid3Δ* had 79 proteins UR and 18 proteins DR. For the

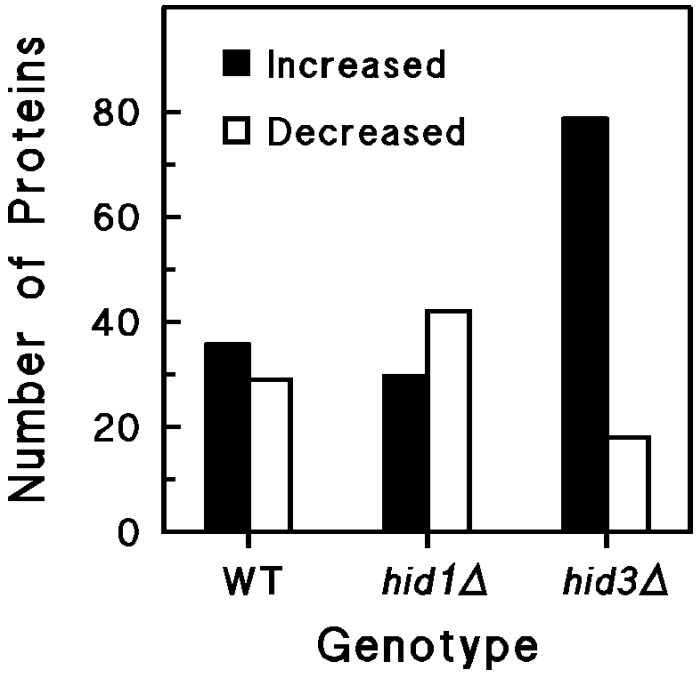
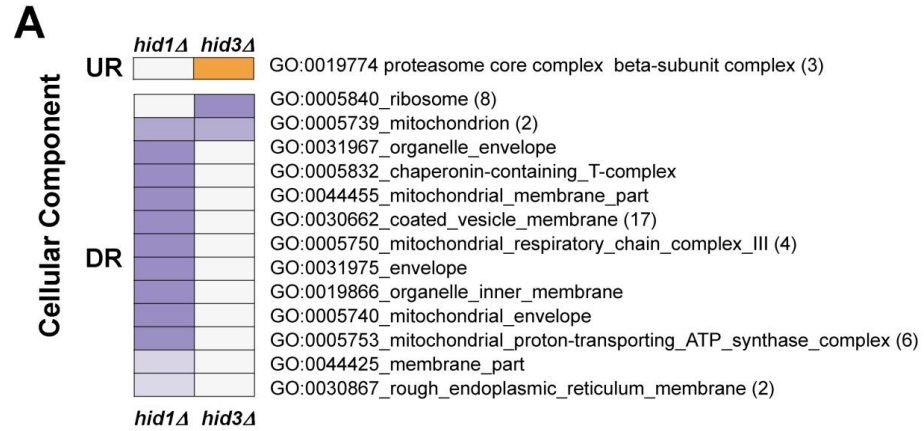
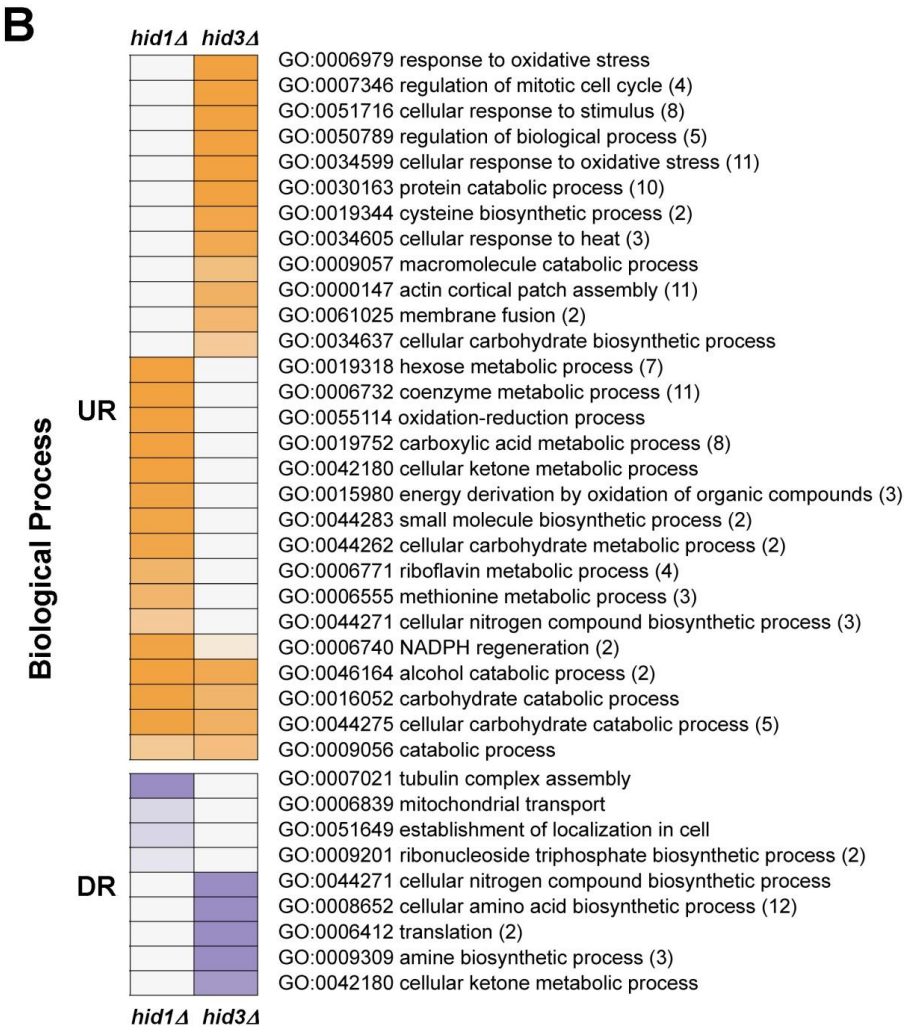


Figure 5.5. The numbers of proteins significantly altered in amount. The numbers of proteins changing were determined by comparison to the VCN genotype. The numbers represent proteins changed in amount by more or less than 2-fold. The determination of relative protein amount was done using Progenesis LC-MS v 4.0 for proteomics (nonlinear DYNAMICS). Progenesis calculated statistical significance using ANOVA with a P-value cut-off of < 0.5.


A
hid1Δ hid3Δ
UR
DR
Cellular Component
GO:0019774 proteasome core complex beta-subunit complex (3)
GO:0005840_ribosome (8)
GO:0005739_mitochondrion (2)
GO:0031967_organelle_envelope
GO:0005832_chaperonin-containing_T-complex
GO:0044455_mitochondrial_membrane_part
GO:0030662_coated_vesicle_membrane (17)
GO:0005750_mitochondrial_respiratory_chain_complex_III (4)
GO:0031975_envelope
GO:0019866_organelle_inner_membrane
GO:0005740_mitochondrial_envelope
GO:0005753_mitochondrial_proton-transporting_ATP_synthase_complex (6)
GO:0044425_membrane_part
GO:0030867_rough_endoplasmic_reticulum_membrane (2)
hid1Δ hid3Δ


B
hid1Δ hid3Δ
UR
DR
Biological Process
GO:0006979 response to oxidative stress
GO:0007346 regulation of mitotic cell cycle (4)
GO:0051716 cellular response to stimulus (8)
GO:0050789 regulation of biological process (5)
GO:0034599 cellular response to oxidative stress (11)
GO:0030163 protein catabolic process (10)
GO:0019344 cysteine biosynthetic process (2)
GO:0034605 cellular response to heat (3)
GO:0009057 macromolecule catabolic process
GO:0000147 actin cortical patch assembly (11)
GO:0061025 membrane fusion (2)
GO:0034637 cellular carbohydrate biosynthetic process
GO:0019318 hexose metabolic process (7)
GO:0006732 coenzyme metabolic process (11)
GO:0055114 oxidation-reduction process
GO:0019752 carboxylic acid metabolic process (8)
GO:0042180 cellular ketone metabolic process
GO:0015980 energy derivation by oxidation of organic compounds (3)
GO:0044283 small molecule biosynthetic process (2)
GO:0044262 cellular carbohydrate metabolic process (2)
GO:0006771 riboflavin metabolic process (4)
GO:0006555 methionine metabolic process (3)
GO:0044271 cellular nitrogen compound biosynthetic process (3)
GO:0006740 NADPH regeneration (2)
GO:0046164 alcohol catabolic process (2)
GO:0016052 carbohydrate catabolic process
GO:0044275 cellular carbohydrate catabolic process (5)
GO:0009056 catabolic process
GO:0007021 tubulin complex assembly
GO:0006839 mitochondrial transport
GO:0051649 establishment of localization in cell
GO:0009201 ribonucleoside triphosphate biosynthetic process (2)
GO:0044271 cellular nitrogen compound biosynthetic process
GO:0008652 cellular amino acid biosynthetic process (12)
GO:0006412 translation (2)
GO:0009309 amine biosynthetic process (3)
GO:0042180 cellular ketone metabolic process
hid1Δ hid3Δ

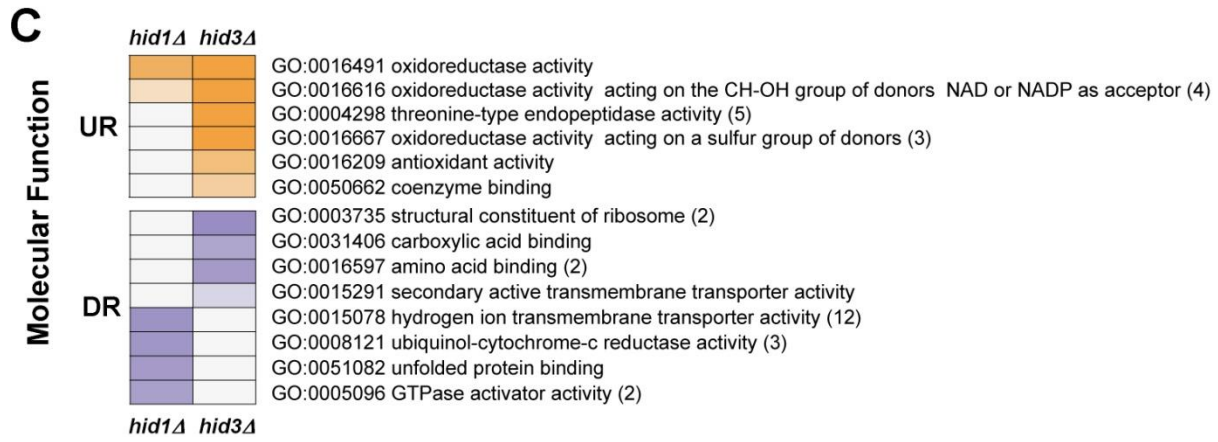


Figure 5.6. Ontology analysis of proteins with significantly altered amounts. The overrepresented classes were determined by GoMiner for significantly changing proteins compared to the set of all proteins used for the normalization set. The UniProt identifiers for all proteins were first converted to the unique S. pombe gene systematic Ids before being run through GoMiner. UR and DR stand for more and less abundant, respectively, than compared to VCN. The GO terms were reduced using REVIGO (revigo.irb.hr, (Supek et al., 2011) and the numbers in parentheses show the number of GO terms combined for the GO term shown. The degree of colour shows the relative degree of statistical significance.

purposes of simplicity WT is also shown compared to VCN (Figure 5.5). Therefore, in order to get an idea on the changes to the proteome due to marker gene expression, it is necessary to reverse the UR and DR proteins of which there were 28 proteins UR and 38 proteins DR compared to WT. The greatest UR protein at 8.3-fold was the pdp3-interacting factor ptf1, which may be involved in chromatin remodelling by affecting the phosphorylation state of Pdp3. This suggests that expression of the marker may affect gene expression, and some minor gene expression differences have been reported in strains expressing the Nat^R gene (Alshehri 2015).

A GO analysis was conducted on the differentially expressed proteins changing by more than 2-fold, however, the numbers of proteins changing were not numerous enough to give statistically relevant enrichment of gene classes, except for a few for *hid3Δ*. Therefore, the GO analysis was repeated but with a set of proteins changing by more than 1.5-fold. For *hid1*Δ there were 134 proteins > 1.5-fold increased and 144 proteins > 1.5-fold decreased compared to VCN. In *hid3Δ*, there were 291 and 347 proteins 1.5-fold UR and DR. Taking the cut-off at 1.5 revealed that approximately the same number of proteins were UR and DR in *hid3Δ.* The only cellular component enriched for either mutant was the beta-subunit complex of the proteosome core-complex. The set of DR proteins for *hid3Δ* were enriched in ribosome and mitochondrial proteins as might be expected for cells under stress conditions. The classes of proteins enriched in the DR set for *hid1*Δ were much more variable, but contained elements mainly involved membrane process associated with the mitochondria. For the biological processes, the UR proteins for *hid1*Δ and *hid3*Δ were divided clearly between metabolic and stress-response processes, respectively. Classes of proteins related to carbohydrate catabolism were also enriched in UR set of proteins for *hid3*Δ. The four classes of that were DR in *hid1*Δ for the biological processes weren't sufficiently coherent to provide an explanation for their enrichment, however, the classes of DR proteins for *hid3*Δ were clearly related to biosynthetic processes that would be reduced in cells under stress, with amino acid and nitrogen metabolism being prominent. In the category Molecular Function the classes of enriched proteins for *hid3*Δ both UR and DR groups could be related to stress responses. The UR proteins were highly represented by oxidoreductases, whereas the DR classes were represented by ribosome, primary metabolic proteins and also transport proteins. The enriched classes of DR proteins for *hid1*Δ appeared to reinforce altered membrane function as initially revealed by the Cellular Component GOs.

Table 5.4. List of proteins in *hid1*Δ with significantly altered amounts. Those proteins are shown that were significantly changed in abundance 2-fold UR or DR in the mutant than in the WT. Proteins were identified using the SEQUEST algorithm in Proteome Discoverer™ (Thermo Scientific).

UniProt Accession	Protein Name	Ratio *hid1*Δ/VCN	Description
2-fold UR			
Q9US41	Urg1	5.3	Uracil-regulated protein
Q9US43	Urg2	3.3	Putative uracil phosphoribosyltransferase
P78831	Ght5	2.8	High-affinity glucose transporter
O94505	Tol1	2.7	3'(2'),5'-bisphosphate nucleotidase
O60181	SPBC23E6.06c	2.7	3,4-dihydroxy-2-butanone 4-phosphate synthase
Q9HEQ9	Tcg1	2.6	Single-stranded TG1-3 DNA-binding protein
O14351	SPBC30D10.05c	2.5	Uncharacterized oxidoreductase
Q9P7F8	Get3	2.4	ATPase
Q09717	Sbp1	2.4	Ran-specific GTPase-activating protein
O14293	SPAC9E9.09c	2.4	Putative aldehyde dehydrogenase-like
Q9US39	SPAC1039.02	2.3	Uncharacterized protein
Q9P782	SPBC1711.08	2.3	Uncharacterized protein
Q9UTJ7	Sdh1	2.3	Probable succinate dehydrogenase,mitochondrial
O42699	Rpl19b	2.3	60S ribosomal protein L19-B
Q09816	SPAC16C9.02c	2.3	S-methyl-5'-thioadenosine phosphorylase
Q9P7B4	SPAC521.03	2.3	Uncharacterized oxidoreductase
P31406	Vma1	2.3	V-type proton ATPase catalytic subunit A
O75000	L12	2.3	60S ribosomal protein L12
P54874	Hcs1	2.3	Hydroxymethylglutaryl-CoA synthase
O36032	Grx1	2.2	Glutaredoxin-1
Q9P7D4	SPBP4H10.15	2.2	Probable aconitate hydratase 2
O74178	Kes1	2.2	Protein kes1
O13991	SPAC26H5.09c	2.2	Uncharacterized oxidoreductase
Q09184	Cdb4	2.2	Curved DNA-binding protein
P31317	Arg3	2.1	Ornithine carbamoyltransferase,mitochondrial
O74893	Rps20	2.1	40S ribosomal protein S20
P87144	Ths1	2.1	Threonine--tRNA ligase, cytoplasmic
Q9P7M9	Rib1	2.1	Probable GTP cyclohydrolase-2
Q9UT27	Pvg1	2.0	Pyruvyl transferase
Q9C0U6	Xks1	2.0	Xylulose kinase
O94373	Atp4	2.0	ATP synthase subunit 4, mitochondrial
2-fold DR			
O94582	Trp3	9.1	Probable anthranilate synthase component 1
Q9Y7U0	SPCC63.14	8.6	Uncharacterized protein
O74440	Rpn8	8.4	26S proteasome regulatory subunit
O94601	SPCC622.14	5.8	Uncharacterized protein
Q9P6R6	Atp16	4.6	ATP synthase subunit delta, mitochondrial
O42940	Pho88	4.2	Inorganic phosphate transport protein
O74417	SPCC14G10.04	4.0	Uncharacterized protein
Q09822	Cdc15	3.7	Cell division control protein 15
O94258	Los1	3.2	Exportin-T
O43024	SPBC354.10	2.9	CUE domain-containing protein
Q9C0W7	Apl3	2.9	AP-2 complex subunit alpha
O14192	SPAC56E4.03	2.9	Aromatic amino acid aminotransferase
O74455	SPCC16C4.10	2.8	Probable 6-phosphogluconolactonase

O14310	Npp106	2.8	Nucleoporin npp106
O74504	SPCC594.01	2.7	UPF0590 protein
Q9UUM7	Hob3	2.7	Protein hob3 (binds cdc42)
Q10367	Glo3	2.5	ADP-ribosylation factor GTPase-activating protein
Q9HGM5	Dbp8	2.5	ATP-dependent RNA helicase dbp8
Q10234	Mrps5	2.5	Probable 37S ribosomal protein S5, mitochondrial
P50999	Cct4	2.4	T-complex protein 1 subunit delta
Q09915	Spt6	2.4	Transcription elongation factor
Q10195	SPBC11C11.06c	2.3	Uncharacterized protein
P87242	SPCC4G3.17	2.3	HD domain-containing protein
O94260	Nxt3	2.3	Putative G3BP-like protein
Q9USP0	Tif221	2.3	Translation initiation factor eIF-2B subunit alpha
O74476	Sal3	2.2	Importin subunit beta-3
O94497	SPBC18E5.07	2.2	Uncharacterized serine-rich protein
O43032	Srp101	2.2	Signal recognition particle receptor subunit alpha homolog
P22068	Atp2	2.2	ATP synthase subunit beta, mitochondrial
O60179	Utp10	2.2	U3 small nucleolar RNA-associated protein 10
O14319	SPBC16E9.02c	2.2	CUE domain-containing protein 5
O94362	SPBC428.15	2.2	Uncharacterized GTP-binding protein
Q9US25	Hrp1	2.2	Chromodomain helicase hrp1
P08091	Pho1	2.2	Acid phosphatase
O59725	SPBC3E7.05c	2.2	Formation of crista-junctions protein
Q9UT11	SPAP8A3.03	2.1	Uncharacterized zinc transporter
O59809	SPCC550.11	2.1	Probable importin
O42932	Qcr6	2.1	Cytochrome b-c1 complex subunit 6
O14369	Sce3	2.1	Probable RNA-binding protein
Q09724	Mrpl11	2.1	54S ribosomal protein L11, mitochondrial
O74874	Ccr4	2.0	Glucose-repressible alcohol dehydrogenase transcriptional effector

Table 5.5. Proteins significantly altered in abundance in *hid3*Δ. Shown are those proteins more (UR) and less abundant (DR) in the mutant vs WT at a level of 2-fold or greater. Proteins were identified using the Proteome Discoverer™ (Thermo Scientific) and relative quantities were determined using the Progenesis QI™ proteome quantification software (nonlinear DYNAMICS).

UniProt Accession	Protein Name	Ratio *hid3*Δ/VCN	Description
2-fold UR			
Q7Z9I3	SPCC663.08c	26.6	short chain dehydrogenase
Q7Z9I4	SPCC663.06c	15.1	short chain dehydrogenase
O94467	SPBC23G7.10c	14.9	Putative NADPH dehydrogenase
O59827	Gst2	5.4	Glutathione S-transferase 2
Q9Y876	Psh3	5.2	ER chaperone SHR3 homologue Psh3
Q09720	SPAC31A2.04c	5.1	Probable proteasome subunit beta type-4
P04913	Htb1	4.3	Histone H2B-alpha
P36627	Byr3	4.1	Cellular nucleic acid-binding protein homolog
O74409	SPCC1223.09	3.3	Uricase
O74875	SPCC330.03c	3.3	Uncharacterized heme-binding protein
P08647	Ras1	3.2	Ras-like protein 1
O74440	Rpn8	3.2	26S proteasome regulatory subunit
O14224	SPAC6F12.06	3.2	Rho GDP-dissociation inhibitor
Q9Y7U0	Eis1	3.2	eisosome assembly protein eis1
O60071	Enp1	3.1	bystin family U3 and U14 snoRNA associated protein (predicted)
O42869	Spa2	3.1	cell polarity protein Spa2
Q9Y7K0	SPBC216.03	3.0	UPF0659 protein
O94312	Pdp2	2.9	PWWP domain-containing protein 2
Q10367	Glo3	2.9	ADP-ribosylation factor GTPase-activating protein
O94246	Gcs2	2.8	glutamate-cysteine ligase regulatory subunit (predicted)
Q10172	Pan1	2.8	Actin cytoskeleton-regulatory complex protein
O74135	Cki3	2.7	Casein kinase I homolog 3
P41891	Gar2	2.7	nucleolar protein required for rRNA processing
Q92357	Cfr1	2.7	Chs five related protein Cfr1
O14295	Plr1	2.7	Pyridoxal reductase
O43021	SPBC354.07c	2.6	Oxysterol-binding protein homolog
O14368	Hsp16	2.6	Heat shock protein 16
O60184	Ssn6	2.6	General transcriptional corepressor
O13791	Slt1	2.6	Schizosaccharomyces specific protein
Q10488	Etr1	2.5	Probable trans-2-enoyl-CoA reductase, mitochondrial
O42940	Pho88	2.5	Inorganic phosphate transport protein
P30821	Obr1	2.5	ubiquitinated histone-like protein Uhp1
O94315	SPBC215.11c	2.5	Uncharacterized oxidoreductase
Q09859	Mxr1	2.4	Probable peptide methionine sulfoxide Reductase
O94601	SPCC622.14	2.4	GTPase activating protein (predicted)
Q02592	Hmt1	2.4	vacuolar transmembrane transporter Hmt1
P41820	Bfr1	2.3	brefeldin A transmembrane transporter
Q9Y7N9	SPCC1450.12	2.3	PX domain-containing protein

O60181	Rib3	2.3	3,4-dihydroxy-2-butanone 4-phosphate synthase
O36032	Grx1	2.3	Glutaredoxin-1
Q10235	Alp11	2.3	tubulin specific chaperone cofactor B
O74401	Ero11	2.3	ER oxidoreductin Ero1a
Q9UUK0	SPAC1952.08c	2.3	Pyridoxamine 5'-phosphate oxidase
P87216	Vip1	2.3	RNA-binding protein Vip1
Q09179	Gln1	2.3	Glutamine synthetase
Q9P7E8	Abp1	2.3	cofilin/tropomyosin family, drebrin ortholog Abp1
Q09770	SPAC1296.01c	2.3	phosphoacetylglucosamine mutase (predicted)
Q9USI5	Sti1	2.2	chaperone activator Sti1 (predicted)
Q9Y7P2	SPCC1450.15	2.2	pig-F/3-ketosphinganine reductase fusion protein (predicted)
Q9USU7	Hmo1	2.2	HMG box protein Hmo1
P04551	Cdc2	2.2	cyclin-dependent protein kinase Cdk1/Cdc2
Q9HGM3	Yta12	2.2	mitochondrial m-AAA protease Yta12 (predicted)
P18253	Ppi1	2.2	cyclophilin family peptidyl-prolyl cis-trans isomerase
Q9C0Y6	SPAPB24D3.08c	2.2	NADP-dependent oxidoreductase (predicted)
Q09745	SPBC12C2.04	2.2	NAD binding dehydrogenase family protein
P17609	Ypt2	2.1	GTPase
P78774	Arc1	2.1	ARP2/3 actin-organizing complex subunit Sop2
Q10169	Dph1	2.1	UBA domain protein
O94579	Pre8	2.1	Probable 20S proteasome complex Subunit alpha 2
Q10494	SPAC26F1.07	2.1	glucose 1-dehydrogenase (NADP+) (predicted)
P33886	Wis1	2.1	Protein kinase
O94497	SPBC18E5.07	2.1	DUF3210 family protein
Q9UUK7	Ade5	2.1	Phosphoribosylglycinamide formyltransferase
Q9P7F4	SPAC2E1P3.01	2.1	Zinc-type alcohol dehydrogenase-like protein
O43026	Gpd3	2.1	Glyceraldehyde-3-phosphate Dehydrogenase 2
P36581	Cnx1	2.1	Calnexin homolog
Q9P7X8	Rpa2	2.1	Probable DNA-directed RNA polymerase I subunit
Q9UU99	SPCC23B6.04c	2.1	sec14 cytosolic factor family (predicted)
O13804	SPAC17H9.07	2.1	Signal recognition particle subunit Srp21 (predicted)
O94733	SPCC191.01	2.0	Schizosaccharomyces specific protein
Q9P3A9	SPAC1565.05	2.0	t-UTP complex subunit Utp8(predicted)
Q9UUM7	Hob3	2.0	BAR adaptor protein Hob3
Q10482	Stm1	2.0	G-protein coupled receptor
O42909	SPBC16A3.02c	2.0	mitochondrial conserved protein (predicted)
O42644	Cts1	2.0	CTP synthase 1
P36586	Ypt5	2.0	GTPase

O13687	Vma4	2.0	V-type ATPase V1 subunit E(predicted)
O60176	Cxr1	2.0	mRNA processing factor
Q10436	SPAC12B10.02c	2.0	ER protein required for packaging into COPII vesicles (predicted)

2-fold DR

Q09852	SPAC23D3.12	9.8	Putative inorganic phosphate transporter
P78831	Ght5	4.0	hexose transmembrane transporter
Q9HE15	SPAC1399.04c	3.8	uracil phosphoribosyltransferase (predicted)
Q9US41	Urg1	3.7	GTP cyclohydrolase II (predicted)
Q9UT95	SPAC323.04	3.6	ABC transporter-like ATPase (predicted)
Q9US43	Urg2	3.1	uracil phosphoribosyltransferase (predicted)
Q8NIL3	Aes1	3.1	enhancer of RNA-mediated gene silencing
P40386	Thi4	2.7	bifunctional thiamine-phosphate dipyrophosphorylase/ hydroxyethylthiazole kinase
O43012	Dcd1	2.5	deoxycytidylate deaminase(predicted)
Q4LB35	fol1	2.4	trifunctional dihydropteroatesynthase/2 amino-4-hydroxy-6-hydro xymethyldihydro- pteridinediphosphokinase / dihydroneopterin aldolase (predicted)
O42885	Pho84	2.4	inorganic phosphate transmembrane transporter (predicted)
O94248	SPCC737.08	2.3	midasin (predicted)
P17478	Rpp202	2.3	60S acidic ribosomal protein A4
O74511	Ekc1	2.3	protein phosphatase regulatory subunit(predicted)
O94377	Atp17	2.2	F0-ATPase subunit F (predicted
Q9USR7	Rpl3801	2.1	60S ribosomal protein L38 (predicted)
O74767	Sec28	2.1	coatomer epsilon subunit (predicted)
P58234	Rps1902	2.0	40S ribosomal protein S19 (predicted)

The lists of proteins that were UR and DR at 2-fold or more are shown for *hid1Δ* and *hid3Δ* are shown in Tables 5.4 and 5.5, respectively. In agreement with the GO analysis inspection of the list of proteins significantly UR in *hid1Δ* does not provide any direct indication of specific processes that were affected. The top two induced proteins were related to uracil metabolism and may reflect the requirement to add supplements to the growth media. The list of DR proteins contains the GTPase activating factor Glo3 and two proteins involved in microtubule formation at the plasma membrane, Cdc15 and Hob3. In contrast, the list of UR proteins for *hid3Δ* reveals increased stress responses and signalling processes related to stress. The top four proteins comprised of the two short-chain dehydrogenases SPCC663.08c (26.6-fold) SPCC663.06c (15.1-fold), a putative NADH dehydrogenase (14.9-fold) and Gst2 (5.4-fold) are induced in response to oxidative (Chen et al., 2003) and/or iron starvation stress (Gabrielli et al., 2012). Other stress-induced proteins listed are uricase, the pyridoxal reductase Plr1 and the heat-shock proteion Hsp16, all which in induced in response to oxidative stress (Chen et al., 2003). GTPase signalling proteins were well represented on the list by the Rho dissociation inhibitor SPAC6F12.06, Ypt2, and many directly related to vesicular transport, such as Ras1, Glo3, Hob3, Stm1 and Ypt5. In addition, there were some GTPase-signaling related proteins, such as gap1, that were induced in the range of 1.5 to 2-fold. Proteins involved in other signaling processes were also induced, such as the cytosolic translational activator Byr3, the transcriptional repressor Ssn6, the hsp-interacting protein Sti1, the cyclin-dependent protein kinase Cdc2, the MAP kinase kinase Wis1 and glyceraldehydes 3-dehydrogenase Gpd3. Other signaling proteins that were increased are associated with vesicular transport, such as the casein kinase Cki3, the Chs five related protein Cfr1, the calnexin homolog Cnx1, and the TF Hmo1. Members of two interesting classes of proteins were induced as well and these have to do with chromatin remodeling and microtubule/cytoskeleton organization. The chromatin structure/remodeling proteins were the histone H2B-alpha subunit Htb1, the PWWP domain-containing protein Pdp2, and the ubiquitinated histone-like protein Uhp1 (Obr1). There were 7 proteins UR that are involved in microtubule/cytoskeleton organization, Abp1, Arc1; Dph1; Hob3; Pan1; Slt1 and Spa2. The list of DR proteins in *hid3Δ* reinforced the decrease in metabolic processes shown by the GO analysis. Of the 347 proteins DR by more than 1.5-fold in *hid3Δ*, 34 were ribosomal proteins or proteins involved in ribosome biogenesis. It also suggested that transport processes that could be affected.

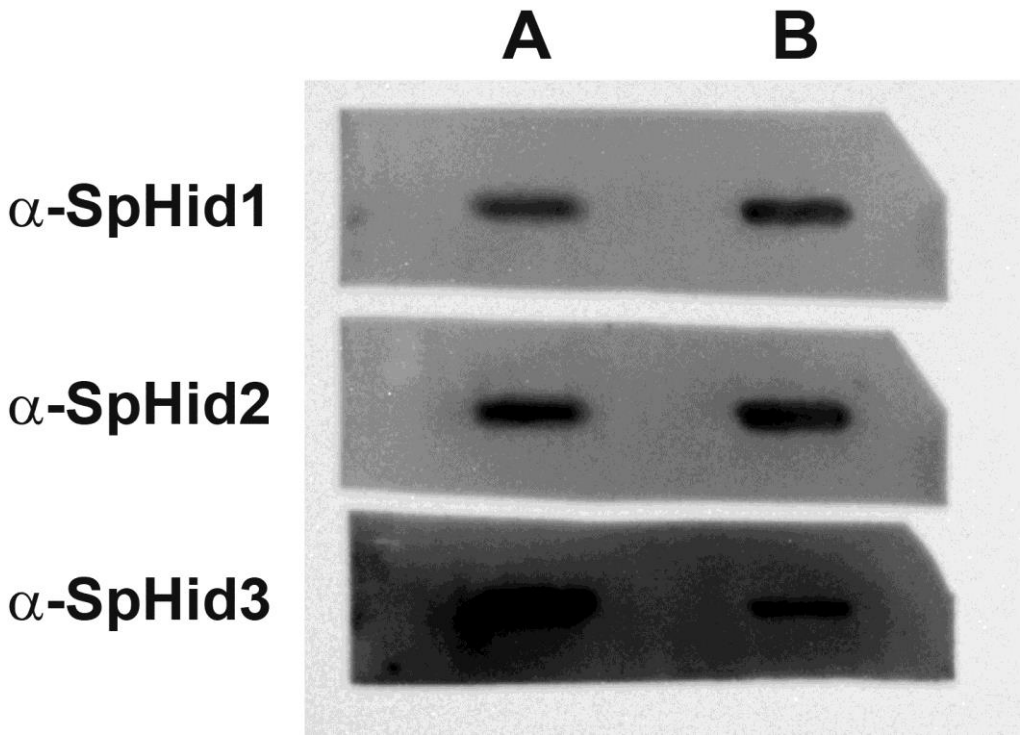


Figure 5.7. Slot blot immunolabeling of *SpHid* proteins in protein extracts of wild-type cells. Cells were grown in standard YES as described in Materials and Methods. Antibodies were generated in Rabbits by PerBio using peptides designed to yield antibodies exhibiting minimal cross-reactivity. The positive control is the antigen peptide for each protein to test cross-reactivity. Labeling and detection was conducted using Invitrogen's WesternDot™ 625 Western Blot Kit according to manufacturers instructions.

5.2.4. Preliminary analyses of immune-detection of Hid proteins in S. pombe

Targeted antibodies to each of the Hid proteins were produced by Beere Antibodies, a subsidiary of Thermo Scientific. First, custom peptides corresponding to the extreme C-terminal sequences of each of the *S. pombe* Hids were synthesized chemically. The sequences used were KREGPGIFEGTDVKIFDSF for anti-Hid1, RQPTILQSRSLGEFQGRLFS for anti-Hid2 and DLLYLQLPSSVNHDSSLRNK for anti-Hid3. Antibodies were produced in rabbits using the Pierce™ 90-day protocol, which starts with an initial injection of 250 µg of peptide antigen in Freund's adjuvant and subsequent booster injections over the 90 day period with 100 µg of antigen. Antibodies from the first bleed were affinity purified using a column with peptide attached. The antibodies were delivered in purified form and as a crude serum along with a sample of the peptide for control blotting.

In order to test the antibody, I used a rapid slot-blotting technique whereby a sample of whole extract and dissolved peptide were deposited onto a PVDF membrane using a membrane holder and vacuum pump. The advantage of this system is that is allowed for rapid testing of the antibody binding in the short time available to work with it. However, it has the disadvantage that it is not possible to distinguish between true signal and background signal, because the proteins were not separated by gel. The slot-blot hybridization is shown in Figure 5.7. For all three antibodies, a signal was detected for crude extract as well as for the purified protein. The signal for Hid1 and Hid2 appeared to be equal to or slightly less than the peptide control. The signal for Hid3 was well above that for the peptide control, thus it appeared that Hid3 is much more abundant than the other two Hids. This indication must be taken with caution, because the relative antibody binding efficiencies and antibody specificities remain to be determined. These must be determined using serial dilutions of peptide and standard Western blots, respectively. One piece of information that supports this finding is that Hid3 was the only Hid paralogue that was detected in the proteomic study being present in one WT and one VCN sample.

5.3 DISCUSSION

This chapter presents the metabolomic and proteomic analyses of the *hid1Δ* and *hid3Δ* mutants. It was not necessarily expected that the metabolite analysis would give any indication of substantial differences in metabolism among genotypes. The priority was the analysis of the proteome and transcriptome. Therefore, in order to minimize changes to the cells, they were immediately frozen after centrifugation without prior washing with water. Even though they were not washed, rapid freezing without water washed would ensure that metabolic changes were minimized. Although on a dry weight basis, the nutrients coming from the yeast extract may not

be detectable by the NMR, but it is possible that the supplements would be detected as adenine-like and UDP-like substances. Nevertheless, the metabolite quantification provided some interesting results. The changes in metabolites we observe were not as dramatic as observed in the null mutant lacking Sty1 or in WT cells treated with H_2O_2 (Weeks et al., 2006), but the directions of metabolite changes corresponded with those of H_2O_2 stress in WT with more glucose, glycerol, acetate, alanine, and lower formate, phenylalanine succinate, valine and leucine. Where the changes didn't match, it was in most cases that we did not observe a change in that particular metabolite, such as for lysine, choline, NADH (nicotinamide) and tyrosine. This would be expected if the stress from the lack of Hid3 was not very strong. From the stress tests, we observe that H_2O_2 has a dramatic negative effect on the growth of *hid3Δ*, thus it would be interesting to check metabolite levels in the mutant under limited H_2O_2 stress. The study of (Weeks et al., 2006) had metabolite assignments in addition to ours, such as for orotic acid, trimethylamine-N-oxide (TMAO), phosphocholine, glycerophosphocholine, and inosine, but we had detected trehalose and Ac-choline, which have also been detected by mass spectrometry (Chaleckis et al., 2014). Some of their novel compounds may be our unknowns or their novel compounds, inosine (adn-like) and UTP + UDP (UDP-like), may be those that we cannot exclude as being from the supplements, like adenine and uracil.

Glycerol, betaine and choline are likely osmoprotective metabolites and their increase in synthesis was hypothesized to represent a cross-over between oxidative and osmotic stress (Weeks et al., 2006). Since we observed a substantial increase in glycerol this suggests that *hid3Δ* cells are under stress. It is difficult to narrow down a particular explanation, since choline was only slightly increased and betaine decreased. Possible explanations for my observations are that the cellular stress caused by the loss of Hid3 is 1) not as severe as direct application of H_2O_2 as mentioned above, 2) only specific stress responses are initiated, or 3) that the chronic stress over generations is somehow attenuated by certain metabolic changes. The latter possibility is supported by the suggestion that for fungi betaine does not serve as an osmoprotectant, but serves as an intermediate in the catabolism of choline (Lambou et al., 2013). Therefore, cells may be undergoing stress conditions without the consequence of synthesizing betaine. These possibilities can be explored by using cell synchronization techniques along with the application of stresses in the mutant genotypes.

In Chapter 3, I had mentioned that potential physical interactors of Hid3 were the ATP-dependent RNA helicase Prp22, and the predicted inositol polyphosphate phosphatase SPAC9G1.10c. In the proteomic analysis, I did not see any change in Prp22 protein, but the ATP-dependent helicase Has1 was significantly UR. This caused me to look a bit at the other

ATP-dependent RNA helicases at both the protein and mRNA level. The gene *has1*+ was significantly UR by 1.7-fold as well as the ATP-depdendent RNA helicase *dbp2*+. Dbp2 protein was also UR suggesting that along with Has1, it is transcriptionally controlled. Two other ATP-dependent RNA helicases had UR mRNA levels, *prh1*+ and *prp43*+. No ATP-dependent RNA helicases were significantly DR at either the protein or RNA level. This result is in direct contrast with the result of (Chen et al., 2003), who showed that all of the ATP-dependent RNA helicases mentioned above are DR in response to all stresses tested. This suggests that the stress response exhibited by removing Hid3 function is fundamentally different to a typical environmentally imposed stress. It would be interesting to see how these genes change in response to oxidative stress in the *hid3Δ*. The putative inositol polyphosphate phosphatase SPAC9G1.10c remained unchanged on both the protein and mRNA level and so its association with Hid3 remains unclear.

The study of (Weeks et al., 2006) also investigated the changing proteome in response to oxidative stress. All but one of the proteins that they observed as being dependent on Sty1 in H_2O_2 induction I also observed as being significantly increased in *hid3Δ*, such as Gst2, SPBC215.11c, Obr1, catalase (1.2-fold induced), and thioredoxin reductase (1.5-fold). This result in itself indicates that *hid3Δ* cells are stressed. This was also the conclusion drawn from transcriptional profiling of *hid3Δ* cells (Alshehri 2015). In addition, I did not see induced any of the proteins that are de-repressed in the unstressed *sty1Δ* mutant. I did observe the induction of Gpd1, which is induced in response to osmotic stress (Ohmiya et al., 1995) and which may be responsible for the production of glycerol observed in the metabolite profiling.

The results mentioned above strongly suggest that Sty1 is active in *hid3Δ* cells. This result is important, because I observed the increase at the protein level of two key components of the MAPK stress-signaling pathway, the Mcs4 response regulator and the MAPKK Wis1, but no induction of Sty1 protein. Mcs4 forms a complex with Wis4, Win1 and Tdh1 (glyceraldehyde 3-phosphate dehydrogenase) in order to mediate the stress response through Wis1. In our study, Tdh1 is reduced 1.2-fold and Wis4 and Win1 do not change on either the protein or gene expression level. However, we see a 2-fold induction in Gpd3, which has been shown to be able to compensate partially for the loss of Tdh1 (Morigasaki et al., 2008). Tdh1 is very abundant compared to the other complex members (10-fold over Gpd3 and 1500-fold over Mcs4 in our study based on raw counts) so that a small reduction of 1.2-fold would unlikely compromise its contribution to the complex. The induction of Wis1 may be sufficient to activate Sty1. Wis1 is necessary for the transition of proliferating to quiescent cell types(Sajiki et al., 2009). I observed some of the hallmarks of the quiescence transition, albeit not on such a dramatic level. These

include the decrease and increase, respectively, in Tdh1 and Gpd3. However, mRNA levels of both genes increase in *hid3Δ*, whereas they decrease in the transition from growth state to quiescent state (Marguerat et al., 2012). I also observed the induction of the GTPase Ypt5, which is required for cell entry into quiescence like Wis1, and the proteins for a number of genes necessary for maintaining quiescence, such as Cdc2, Nat10, Apl2, and Cts1. Is it possible that the slow growth phenotype is simply a result of elevated Cdc2, which could be phosphorylated and slow cell cycle progression? However, the kinases Chk1 and Cds1 that phosphorylate Cdc2 (Alao and Sunnerhagen, 2008) remained unchanged at both the protein and mRNA level. It is also interesting to speculate that the decreased growth phenotype (Chapter3, Alshehri 2015) is indicative of a mild quiescent state brought on by stress. The main question arising is how is the lack of Hid3 being sensed as a stress.

Chapter 6

General discussion and future work

The experimentation that is presented in this work went in three phases. The first phase was to conduct a bioinformatic analysis of the HIDs in order to establish the ground-work for using *S. pombe* as a model for HID function in animals, particularly humans. The second phase built on this, and on preliminary growth and morphology characterisation of the mutants conducted by Alshehri (2015), as a functional study in order to investigate the stress responses of the *hid* mutants, particularly *hid1* and *hid3*, and the role of Hid3 in the Golgi. The third phase involved a proteomic study to determine if protein levels would change according to the stress responses and various models that had been proposed in the literature to ascribe a function to Hid3. I have divided this discussion into four sections each relating to each of the Chapters of the thesis and a presentation of future areas of work. I hope that is it realised that this discussion couldn't cover all hypotheses or mechanisms that may arise from the experimentation, particularly from the proteomics.

I. Chapter 3 conclusions

From the literature, there were three major ideas regarding the potential function of HID1 in humans. The first one came in 2001 with the publication by (Harada et al., 2001) that deletion or reduction of expression of HID1 (then c17orf28 or DMC1) contributed to cell proliferation in a number of cancer types. The second was showing that HID-1 of *C. elegans* is required for normal vegetative growth of the worm (Ailion and Thomas, 2003). The third came from S. pombe itself showing that elimination of Hid3 (then and still Ftp105) caused mis-localisation of the deubiquitinase Ubp5 from the Golgi to the cytosol (Kouranti et al., 2010). More recently, general screening of S. pombe mutants in relation to quiescence due to nitrogen starvation suggested the involvement of Hid3 in maintaining normal growth (Sideri et al., 2014). Seeing that there are strong similarities between S. pombe and human cell physiology regarding Golgi function, it is possible that S. pombe Hid3 would serve as a good model to address the functions mentioned above.

The computational approach was attempted to identify aspects of HID function that were not already present in the literature, including potential complications for using S. pombe as a model system. The various databases, such as NCBI, Uniprot, etc. referred to HID1 as being conserved from fungi to humans, which is quite small on the evolutionary scale. In NCBI,

because of sequence similarities, HID1 had been assigned to the DYMECLIN superfamily, which is conserved from plants to humans. Although these mentions come from the literature, to my knowledge, there has not been a detailed comparison of DYM and HID1

Structure or evolutionary origins. HID and DYM are both present in animals so knowing their evolutionary origins and structural characteristics may give information on functional overlap. It wasn't possible to uncover a protein that was ancestral to both DYMs and HIDs. They are both common only to eukaryotes and are present in the eukaryotes prior to the division of the seven superphyla (Adl et al., 2005). However, it was interesting to find that whereas animals have both DYM and HID, plants have only DYM and fungi have only HID. Having DYM or HID is not a requirement for a cell being eukaryotic as shown by a vast number of organisms not having DYM or HID. Although, we have not identified a multicellular eukaryote with the exception of species of linked algae, that do not have one or the other. Thus, it is interesting to speculate that having either is necessary for cellular communication that makes multicellularism possible, but it also suggests there may be functional overlap. It is interesting to speculate that DYM is not a lethal mutation in humans, because it could be partially complemented by HID1. DYM represents a possible complication for understanding the complete function of HID1, but it is also interesting to understand how both work within the same context of Golgi function.

The major differences that have been noted between the lineages of the two proteins, such as the HID and DYM present within the choanoflagellates, the presence of DYM within the fungal family Mucorales, and the apparently different evolutionary rates of the HIDs and DYMs within fungi and plants, respectively, directly suggest that horizontal gene transfer took place at some point within their evolutionary histories, which is known to happen between eukaryotes (Keeling and Palmer, 2008). Species of Mucorales are parasitic and they can infect animals, plants or each other. For example, species of absidia can infect animals or plants and species of parasitica can infect and pass or retain genes from the absidia (Hoffmann et al., 2013; Kellner et al., 1993). Because the DYMs are restricted to the mucorales, it is interesting to speculate that they may have obtained them from another eukaryote, in this case from a non-fungal Opisthokont. Far reaching speculation is that they subsequently transferred them to the Viridiplantae, noting that the plant DYMs are more evolutionarily conserved to animals than the HIDs are within fungi.

The annotation of the *S.pombe* genome revealed that there are three orthologues of HID1. Two of the orthologues had been named as human HID1 orthologue 1 (SPAP27G11.12) and HID1 orthologue 2 (SPBP19A11.07c). Reference to the human orthologue HID1 has only

come about within the last year and these had been referred to as orthologues of human DMC1, the term given by (Harada et al., 2001) for the Down-regulated in Multiple Cancers 1 locus. The phylogenetic analysis indicates that in the ascomycetes the ancestral *HID* gene duplicated to become *HID2* and the ancestor of *HID1/HID3*, which then duplicated to become *HID1* and *HID3*. The presence of three paralogues may be unusual, but interesting from the point of redundancy of function. Since Hid1 and Hid3 are the most similar they were the basis of the systems biology study of the group. Why are there three orthologues of HID1 in pombe? One hypothesis is that they split to retain the corresponding (as of yet unknown) functions of HID1 in humans and *C. elegans*. There is evidence that HID1 dynamically associates with the Golgi membrane (Wang et al., 2011). This implies that HID1 has a function when associated with the Golgi membrane and another when released into the cytosol. Is it possible that Hid3 retains the function of Golgi-associated HID1 and Hid1 of that when present in the cytosol and is it possible that the function of Hid3 has been modified to become an integral protein for maintaining Golgi structure in S. pombe (Alshehri ,2015).

II Chapter 4 conclusions

From the work conducted on *C. elegans*, the conclusions were made that HID-1 functions in the transport of neuropeptides in neuronal cells, the type of cells where it is most highly expressed. This is based on the observation that HID-1 is present at the synaptic junctions of the plasma membrane as well as being associated with exocytotic vesicles (Mesa et al., 2011; Yu et al., 2011). Although it cannot be entirely excluded that HID1 works in a retrograde manner to recycle proteins or lipids from the plasma membrane to the endosome or Golgi, it most likely has an anterograde transport role. So far for pombe, Hid3 has only been associated with the Golgi apparatus and so its role in anterograde and/or retrograde protein transport remained unknown. In general, it is difficult to differentiate between anterograde and retrograde transport processes because they are highly interdependent, but it is possible by using certain inhibitors of Golgi function. By using the inhibitors BFA and GCA, I showed that growth of *hid3Δ* was unaffected by low concentrations of BFA, but that of WT and VC genotypes was. At high concentrations of BFA, growth of all genotypes was dramatically slowed. This result was consistent with the broad inhibitory nature of BFA to act on the Golgi and disrupt its structure. Compared to WT, VC and *hid1Δ* genotypes, *hid3Δ* was hypersensitive to the compound GCA, which is known to inhibit retrograde protein transport from endosomes to the Golgi (Sáenz et al., 2009). The broad conclusion that can be made is that in the *hid3Δ* mutant

GCA disrupts those processes that are not affected by the lack of Hid3 and so completely abolishing Golgi function. I would like to speculate that Hid3 functions primarily to maintain anterograde protein transport in *S. pombe*. This is consistent with its subcellular location to the *medial-* and *trans*-Golgi as reported for *C. elegans* (Wang et al., 2011).

The genome wide gene expression data provided by (Chen et al., 2003) gave the suggestion that the Hid proteins may be involved in the stress responses and so should be test for their responses to stress. The genes *hid1*$^{+}$ and *hid2*$^{+}$ were both induced in response to multiple stresses, but *hid3*+ was not. *Hid3*$^{+}$ expression was essentially unchanged suggesting a primary role in maintaining Golgi function even under stress conditions. Over the variety of stresses tested, both abiotic and DNA damaging, *hid1Δ* and *hid3Δ* genotypes were relatively unaffected, except with oxidative stress through treatment with hydrogen peroxide where both were affected. In contrast, it appeared that *hid2Δ* was not. There is an explanation for this similar result with *hid1Δ* and *hid3Δ*. Hid1 may be involved directly in the cellular response to oxidative stress as there are a number of important proteins DR in the mutant (see Chapter 5 below). It is likely that Hid3 minus cells are already compromised in their growth and so an added stress could further reduce it. Maybe *hid3Δ* cells are already sensitised for oxidative stress responses through the constitutive induction of signalling pathways and so cannot further respond to the stress. The *hid3Δ* mutant has been reported to be moderately sensitive to cadmium and high KCl treatment (Kennedy et al., 2008). Although I did not test cadmium sensitivity I did not observe sensitivity to high NaCl. It should be noted that the already reduced growth of *hid3Δ* was a complicating factor in trying to determine sensitivity to stress, because all evaluations had to be done on a relative basis on how mutant growth was affected proportionately to that of WT or VC genotypes. It is possible that some stress treatments may have had effects that were missed.

As mentioned in Chapter 4, it was unexpected that *hid3Δ* would grow very poor in Edinburgh Minimal Medium. This was observed in liquid culture monitored by cell counts as well as by spot size on agar plates using standard drop tests. It is not clear why, but growth was affected depending on the type of EMM used. All genotypes grew poorly in the EMM from Sunrise Science Products and better with the Foremedium EMM. However, it is clear that even Foremedium EMM does not support complete growth even of WT cells so there is some type of nutrient limitation. Since the growth of *hid3Δ* was so much worse than the other genotypes, I can conclude that the nutrient limitation effect is much more profound in *hid3Δ*. In addition, a nutrient limitation effect was supported by experiments where standard yeast extract media was titrated in increasing amount into EMM with the effect to improve the growth of all genotypes,

but improving *hid3Δ* growth required higher proportions of yeast extract. The conclusion of the poor growth phenotype associated with nutrient limitation is consistent with the gene expression results showing induction of gene expression for a variety of plasma membrane nutrient transporters (Alshehri 2015). What has been a big subject of recent research is the switch between quiescent and proliferative cell growth. The poor growth of *hid3Δ* on EMM is consistent with nutrient limitation inducing a quiescent growth state. Naturally, it is essential to determine which element is preventing growth of *hid3Δ* strains.

III Chapter 5 conclusions

Chapter 5 presented the results from the metabolomic and proteomic study of the *hid* mutants. These two technologies were used to try to address some of the current hypotheses about Hid function. The metabolomics would show if any large changes to metabolism occurred in the *hid* mutants, and the proteomics would provide some mechanistic explanation for the slow growth observed for *hid3Δ*. It has been proposed that *hid3Δ* cells may be in a sort of quiescent growth state similar to *hid-1* mutants of C. elegans (Alshehri 2015). A number of genes and metabolic conditions have been described for S. pombe cells to enter and maintain a quiescent growth state (Gray et al., 2004; Sajiki et al., 2009). It would be expected that these properties would be present in *hid3Δ* cells if they were in a quiescent or partially quiescent state.

The metabolomics experiments quantified 27 metabolites in WT, control and *hid1Δ* and *hid3Δ* genotypes. Certainly, the most surprising was the difference in glucose levels between the WT and strains expressing the antibiotic marker. Why this should be the case is not clear. The data showed two interesting differences in metabolite content between *hid3Δ* mutants and VC strains, which were an increase in glycerol and trehalose. An increase in these metabolites, at least in budding yeast, is seen as necessary for cells to maintain the quiescent state. Trehalose is reported to protect quiescent cells from reactive oxygen species (Benaroudj et al., 2001) and glycerol is likely to prevent cell desiccation under long term survival. Although the changes were not dramatic, the *hid3Δ* mutants showed some metabolic changes consistent with a quiescent state.

The label-free proteomics turned out to be an efficient and relatively inexpensive way to determine protein changes associated with the *hid* deletions. I was able to observe the fate of nearly 2000 proteins present in S. pombe cells. The most striking result from the proteomics analysis of the *hid* mutants was the pleiotropic nature of the protein changes in *hid3Δ*. The lists provided in Tables 5.4 and 5.5 show that 72 and 92 proteins were significantly DE in *hid1Δ* and *hid3Δ*, respectively, at a level of 2-fold or greater. This presented somewhat of a narrow picture

of the entire landscape of proteins changes seeing that totals of 467 and 712 proteins were significantly DE in *hid1Δ* and *hid3Δ*, respectively. For each mutant, the complete sets of DE proteins were used in the GO analysis in order to establish the functional classes of proteins that changed in relation to the VC genotype. The proteins changing gave interesting insights into the processes that were changing in both mutants. For *hid1Δ*, organelle-related membrane processes were DR and cytosolic metabolic processes were UR. This suggests a fundamental metabolic shift controlled by loss of Hid1. This did not have any effect on growth rate under standard conditions, but it compromised the mutant's ability to cope with oxidative stress. Mitochondria are highly sensitive to increased reactive oxygen species that disrupt membrane integrity and respiration, and the lack of mitochondrial function is well known to lead to apoptosis (Circu and Aw, 2010). An interesting idea is that Hid1 in S. pombe contributes to maintaining interorganelle communication of metabolic processes and that this may be the function of HID1 dissociation from the Golgi in animals cells.

It has been proposed that Hid3 may sequester Ubp5 to the golgi to prevent it from entering the nucleus to deubiquitinate certain transcription factors. This system would be analgous to the MDM2-DAXX-USP7 complex that appears to regulate the nuclear localisation of both pTEN and p53 and subsequently affects induction of apoptotic pathways. In humans, USP7 deubiquitinates pTEN leading to its translocation to the cytosol and preventing activation of apoptotic pathways (Song et al., 2008). When DNA damage occurs DAXX dissociates from MDM2 to disrupt the complex causing MDM2 to degrade through autoubiquitination, which is prevented by USP7 in non-stress cells. Degradation of MDM2 prevents p53 ubiquitination and enhances it lifespan allowing it to induce cell apoptotic pathways (Zhao et al., 2004). USP7 directly deubiquitinates p53 allowing activation of apoptotic pathways (Li et al., 2002). According to these two models, HID1 reduction would lead to increased availability of USP7 and either promote or prevent cell proliferation depending on the cancer type. In fact, it is quite rare to have both pTEN and p53 mutations in the same cancer and so they serve as backup systems (Li et al., 2008b). These types of models, although unclear, demonstrate that disruption of Golgi function could affect gene expression.

My data provides a few interesting examples of how disrupting Golgi function may affect chromatin structure and gene expression. One example is Midasin, a nuclear protein that facilitates the biogenesis of ribosomes by aiding their export by removing biogenesis factors (Bassler et al., 2010). Midasin has significant sequence similarity to human DAXX. Midasin protein was DR by more than 2-fold in *hid3Δ* and so could be part of the explanation for the

down-regulation of ribosome function in the mutants. Down-regulation of ribosome function and protein synthesis are symptoms of both cells under chronic stress and those entering quiescence. Two interesting proteins that were UR in *hid3Δ* are Htb1 and Obr1/Uhp1. Obr1/Uhp1 has a histone fold similar to Histone H2A and it binds Histone H2B of which Htb1 is a subunit. The current model of Uhp1 function is in chromatin silencing at mating locus *mat2P* through dynamic chromatin interaction regulated by ubiquitination through the E1 ligase Rhp6 (Naresh et al., 2003). Rhp6 was not affected in our study at the level of protein or gene expression. Obr1/Uhp1 is proposed to help maintain a semi-stable structure of chromatin that will then become stable upon Uhp1 ubiquitination and, subsequent, targeting for degradation. The presence of Ubp5 in the cytosol/nucleus may be responsible for maintaining elevated levels of Obr1/Uhp1 by deubiquitinating it. The result would be maintaining an active euchromatin state at the mating switch locus through its interaction with H2B as opposed to H2A. Therefore, a combination of an increase in Obr1/Uhp1 and H2B along with deubiquitination would maintain the mating-type locus in an non-silent state. Induction of Interestingly, activation of pheromone sensing and signalling pathways leading to unsilencing of mating type switching are mediated by Ras signalling when Ras is located at the plasma membrane (Onken et al., 2006; Tamanoi, 2011). I observed more than a three-fold increase in Ras1 in *hid3Δ* mutants.

These examples are speculation as to the effects that Ubp5 may have when Hid3 is deleted. It must be considered that ubiquitination/deubiquitination is a process that occurs in many different cellular compartments including the Golgi apparatus. Ras signalling itself is tuned by ubiquitination/deubiquitination processes (Jura et al., 2006). Within the ER, ubiquitination serves as a signal to export improperly folded proteins to the cytosol for degradation by the proteosome. In contrast, it is believed that deubiquitination serves as a check that properly folded/processed proteins are not targeted for degradation (Zhang et al., 2013a). A very good example is the SREBP signalling processes in lipid metabolism. SREBP is an ER resident protein that is shuttled to the Golgi where it is activated by ubiquitination to subsequently be cleaved for release into the cytosol and transport to the nucleus (Stewart et al., 2011).

There are a number of additional noticeable and interesting features to the set of UR proteins in *hid3Δ*. Discussed quite extensively was the induction of the MAPK stress-response signaling pathway and oxidative stress-response proteins, such as the complement of short-chain dehydrogenases, NADPH dehydrogenases and glutathione metabolism proteins. One point that needs to be added here is that the induction of many of these proteins, such as the short-chain dehydrogenase Osr1 and Wis1 appear to be under transcriptional control (Alshehri 2015) and so demonstrating that *hid3Δ* is under chronic stress and is trying to adapt or cope

with it. One way to cope with chronic stress is either through cell death, which is not really an option for single cell organisms, or to adopt a quiescent state of growth. The metabolomic and protein profiling are consistent with the hypothesis that cells are in a quiescent state (Alshehri 2015). One interesting question is if adoption of a quiescent state is compatible with activation of mating-type switching. Indications are that is it. Nitrogen starvation induces mating-type switching (Sun et al., 2013) as well as quiescence (Sideri et al., 2014). In addition, the induction of autophagy, which may be important in the nutrition of quiescent cells is depend on mating type (Sun et al., 2013). A major contrast of the proteomic results from those of the transcriptics is the lack of UR of transporter proteins, but the genes are UR. Some transporter proteins mainly involving phosphate transport were shown to be DR. This result supplies part of the mechanism to explain the changes in gene expression. The lack of transport capacity leads to starvation of an element (unknown at this time) leading to increased expression of the genes to make more protein. Even if protein is being made under a reduced biosynthetic capacity, it is unlikely that they are reaching the plasma membrane, because of disrupted Golgi function, and so are being degraded.

IV Future work

The phylogeny is an ongoing process that will require more detailed analysis of sequence properties. There are a number of sequences that fell below the criteria for selection within the global phylogenetic analysis that could provide more information as to early common features between the HIDs and DYMs. Also, it would be useful to try to root the trees using a more divergent sequence, such as c10orf76, which is the most similar protein pulled up by the BLAST searches. In an experimental approach is will be necessary to express the human and *C. elegans* HIDs and DYMs in *hid3Δ* in order to see if any complement the poor growth phenotype on minimal media. It has recently been reported that *hid3Δ* strains are hypersensitive to the drug bortezomib (Takeda et al., 2011). The sensitivity to this drug is another way to test for functional complementation with the animal proteins.

It will certainly be necessary to investigate the reason for both *hid1Δ* and *hid3Δ* to be sensitive to oxidative stress. The complementation of the oxidative stress sensitive phenotype in the *hid1Δ hid2Δ* double mutant suggests that it may be possible to look for genetic suppressors in order to identify the genes responsible. Likewise, it will be necessary to determine why EMM has such a negative effect on *hid3Δ* growth. It may be possible to test if the re-addition of any particular component(s) to the minimal media may be able to stimulate *hid3Δ* growth.

From a biochemical aspect it would be interesting to determine the orientation of Hid3 in Golgi membranes and why it is important for maintaining Golgi Structure. There are tools that have been established for plants that can be adapted for use in S. pombe (Søgaard et al., 2012) and the use of the antibodies will be important for these studies. The antibodies may also be useful for identifying proteins that interact with Hid3 as has been done for Upb5 with a TAP-tagged Hid3 (Kouranti et al., 2010). The purpose of showing the antibody results, although very preliminary, is that I am developing some of the tools necessary to investigate protein interactions with Hid3, which I had discussed quite extensively in Chapter 3. In addition, they can be used to investigate the subcellular localization of the Hids, through in immuno hybridization for example.

Besides mentioning Ras1, I have not discussed the dramatic changes to GTPase signaling proteins that I observed. This includes the large number of actin cortical patch assembly proteins that were seen to be UR. First, it will be necessary to determine if their subcellular location, particularly that of Ras1, changes in *hid3Δ*.

A major question is what is regulating protein changes in *hid3Δ*? For cell stress mRNA and protein levels are generally correlated (Lackner et al., 2012) and this is what we see for the limited metabolic genes and proteins related to stress. There are three possible mechanisms to

explain the change in protein levels. One way is through altering transcription to produce more or less mRNA for translation. This can be addressed by looking at levels of directly transcribed mRNA that are integrated into polysomes called Ribosome profiling (Ingolia, 2014). mRNA from polysomes can be quantified using RNAseq and this can be associated with the protein composition (Fuchs et al., 2011). The second way is through mislocalisation of Ubp5 leading to altered ubiquitination in the cytosol and nucleus. Protein turnover under ubiquitination/deubiquitination control involving Ubp5 can be determined by use of the *ubp5* mutant. Finally, proteins that are mislocalised due to disruption of the Golgi complex could be determined by using a brefeldin A to disrupt the Golgi complex. For example, Ras1 is a very good example of a protein that does not show transcriptional regulation of its levels in hid3Δ. The gene expression analysis conducted by Alshehri (2015) showed that *ras1+* transcript levels do not change in the mutant compared to VCs, WT or *hid1Δ* genotypes. Disruption of the Golgi apparatus may remove Ras1 from one of a number of Golgi located D3 ubiquitin ligases, therefore, preventing its ubiquitination and subsequent turnover. Alternatively, disruption of the Golgi may bring Ras1 in more frequent contact in order to be ubiquitinated. This would entail a combined study to identify proteins changing in common and distinctly. This could be expanded to include both oxidative stress signalling and nutrient signalling.

REFERENCES

A, A.M., Maucourt, B., A, C.R., Gaudill, M., C, J.V., Brouquisse, R., D, D.G., Denoyes-rothan, A., Lerceteau-k, E., and B, D.R. (2004). Quantitative metabolic profiling by 1-dimensional 1 H-NMR analyses : application to plant genetics and functional genomics. 889–902.

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., et al. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. *52*, 399–451.

Aguiar, T.Q., Ribeiro, O., Arvas, M., Wiebe, M.G., Penttilä, M., and Domingues, L. (2014). Investigation of protein secretion and secretion stress in Ashbya gossypii. BMC Genomics *15*, 1137.

Ailion, M., and Thomas, J.H. (2003). Isolation and characterization of high-temperature-induced Dauer formation mutants in Caenorhabditis elegans. Genetics *165*, 127–144.

Alao, J.P., and Sunnerhagen, P. (2008). Rad3 and Sty1 function in Schizosaccharomyces pombe: an integrated response to DNA damage and environmental stress? Mol. Microbiol. *68*, 246–254.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular Biology of the Cell (Garland Science).

Alshehri, M. (2015) Characterisation of transcriptional properties of the *hid1Δ* and *hid3Δ* mutants of Schizosaccharomyces pombe. PhD Thesis. University of Bordeaux.

Alshammari, M.J., Al-Otaibi, L., and Alkuraya, F.S. (2012). Mutation in RAB33B, which encodes a regulator of retrograde Golgi transport, defines a second Dyggve–Melchior–Clausen locus. J. Med. Genet. jmedgenet – 2011–100666.

Baldauf, S.L. (2003). The deep roots of eukaryotes. Science *300*, 1703–1706.

Bassler, J., Kallas, M., Pertschy, B., Ulbrich, C., Thoms, M., and Hurt, E. (2010). The AAA-ATPase Rea1 drives removal of biogenesis factors during multiple stages of 60S ribosome assembly. Mol. Cell *38*, 712–721.

Bayrak, I.K., Nural, M.S., and Diren, H.B. (2005). Dyggve-Melchior-Clausen syndrome without mental retardation (Smith-McCort dysplasia). Diagn. Interv. Radiol. Ank. Turk. *11*, 163–165.

Bem, D., Yoshimura, S.-I., Nunes-Bastos, R., Bond, F.C., Bond, F.F., Kurian, M.A., Rahman, F., Handley, M.T.W., Hadzhiev, Y., Masood, I., et al. (2011). Loss-of-function mutations in RAB18 cause Warburg micro syndrome. Am. J. Hum. Genet. *88*, 499–507.

Benaroudj, N., Lee, D.H., and Goldberg, A.L. (2001). Trehalose accumulation during cellular stress protects cells and cellular proteins from damage by oxygen radicals. J. Biol. Chem. *276*, 24261–24267.

Bentley, N.J., Holtzman, D.A., Flaggs, G., Keegan, K.S., DeMaggio, A., Ford, J.C., Hoekstra, M., and Carr, A.M. (1996). The Schizosaccharomyces pombe rad3 checkpoint gene. EMBO J. *15*, 6641–6651.

Bergink, S., and Jentsch, S. (2009). Principles of ubiquitin and SUMO modifications in DNA repair. Nature *458*, 461–467.

Bernard, P., Hardwick, K., and Javerzat, J.P. (1998). Fission yeast bub1 is a mitotic centromere protein essential for the spindle checkpoint and the preservation of correct ploidy through mitosis. J. Cell Biol. *143*, 1775–1787.

Bertram, J.S. (2001). The molecular biology of cancer. Mol. Aspects Med. *21*, 167–223.

Bexiga, M.G., and Simpson, J.C. (2013). Human Diseases Associated with Form and Function of the Golgi Complex. Int. J. Mol. Sci. *14*, 18670–18681..

Birnby, D.A., Link, E.M., Vowels, J.J., Tian, H., Colacurcio, P.L., and Thomas, J.H. (2000). A transmembrane guanylyl cyclase (DAF-11) and Hsp90 (DAF-21) regulate a common set of chemosensory behaviors in caenorhabditis elegans. Genetics *155*, 85–104.

Boissé Lomax, L., Bayly, M.A., Hjalgrim, H., Møller, R.S., Vlaar, A.M., Aaberg, K.M., Marquardt, I., Gandolfo, L.C., Willemsen, M., Kamsteeg, E.-J., et al. (2013). “North Sea” progressive myoclonus epilepsy: phenotype of subjects with GOSR2 mutation. Brain J. Neurol. *136*, 1146–1154.

Bonifacino, J.S., and Glick, B.S. (2004). The Mechanisms of Vesicle Budding and Fusion. Cell *116*, 153–166.

Brinkmann, K., Schell, M., Hoppe, T., and Kashkar, H. (2015). Regulation of the DNA damage response by ubiquitin conjugation. Front. Genet. *6*, 98.

Brown, J.S., and Jackson, S.P. (2015). Ubiquitylation, neddylation and the DNA damage response. Open Biol. *5*, 150018.

Buchberger, A., Bukau, B., and Sommer, T. (2010). Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. Mol. Cell *40*, 238–252.

Burande, C.F., Heuzé, M.L., Lamsoul, I., Monsarrat, B., Uttenweiler-Joseph, S., and Lutz, P.G. (2009). A label-free quantitative proteomics strategy to identify E3 ubiquitin ligase substrates targeted to proteasome degradation. Mol. Cell. Proteomics MCP *8*, 1719–1727.

Carpy, A., Krug, K., Graf, S., Koch, A., Popic, S., Hauf, S., and Macek, B. (2014). Absolute proteome and phosphoproteome dynamics during the cell cycle of Schizosaccharomyces pombe (Fission Yeast). Mol. Cell. Proteomics MCP *13*, 1925–1936.

Chaleckis, R., Ebe, M., Pluskal, T., Murakami, I., Kondoh, H., and Yanagida, M. (2014). Unexpected similarities between the Schizosaccharomyces and human blood metabolomes, and novel human metabolites. Mol. Biosyst. *10*, 2538–2551.

Chang, E.C., and Philips, M.R. (2006). Spatial segregation of Ras signaling: new evidence from fission yeast. Cell Cycle Georget. Tex *5*, 1936–1939.

Chappell, T.G., and Warren, G. (1989). A galactosyltransferase from the fission yeast Schizosaccharomyces pombe. J. Cell Biol. *109*, 2693–2702.

Chen, D., Toone, W.M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N., and Bähler, J. (2003). Global transcriptional responses of fission yeast to environmental stress. Mol. Biol. Cell *14*, 214–229.

Chen, P.-C., Na, C.H., and Peng, J. (2012). Quantitative proteomics to decipher ubiquitin signaling. Amino Acids *43*, 1049–1060.

Chhangani, D., Joshi, A.P., and Mishra, A. (2012). E3 ubiquitin ligases in protein quality control mechanism. Mol. Neurobiol. *45*, 571–585.

Circu, M.L., and Aw, T.Y. (2010). Reactive oxygen species, cellular redox systems, and apoptosis. Free Radic. Biol. Med. *48*, 749–762.

Clague, M.J., Coulson, J.M., and Urbé, S. (2012). Cellular functions of the DUBs. J. Cell Sci. *125*, 277–286.

Coburn, C.M., Mori, I., Ohshima, Y., and Bargmann, C.I. (1998). A cyclic nucleotide-gated channel inhibits sensory axon outgrowth in larval and adult Caenorhabditis elegans: a distinct pathway for maintenance of sensory axon structure. Dev. Camb. Engl. *125*, 249–258.

Connerly, P.L. (2010). How do proteins move through the Golgi apparatus. Nat Educ *3*, 60.

Corbett, M.A., Schwake, M., Bahlo, M., Dibbens, L.M., Lin, M., Gandolfo, L.C., Vears, D.F., O'Sullivan, J.D., Robertson, T., Bayly, M.A., et al. (2011). A mutation in the Golgi Qb-SNARE gene GOSR2 causes progressive myoclonus epilepsy with early ataxia. Am. J. Hum. Genet. *88*, 657–663.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. Science *327*, 425–431.

Costanzo, M., Baryshnikova, A., Myers, C.L., Andrews, B., and Boone, C. (2011). Charting the genetic interaction map of a cell. Curr. Opin. Biotechnol. *22*, 66–74.

Curtis, B.A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M.C., Ball, S.G., Gile, G.H., Hirakawa, Y., et al. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature *492*, 59–65.

Dang, C. V (2012). Links between metabolism and cancer. Genes Dev. *26*, 877–890.

Darling, N.J., and Cook, S.J. (2014). The role of MAPK signalling pathways in the response to endoplasmic reticulum stress. Biochim. Biophys. Acta *1843*, 2150–2163.

Denais, C., Dent, C.L., Southgate, L., Hoyle, J., Dafou, D., Trembath, R.C., and Machado, R.D. (2011). Dymeclin, the gene underlying Dyggve-Melchior-Clausen syndrome, encodes a protein integral to extracellular matrix and golgi organization and is associated with protein secretion pathways critical in bone development. Hum. Mutat. *32*, 231–239.

Dillin, A., Crawford, D.K., and Kenyon, C. (2002). Timing requirements for insulin/IGF-1 signaling in C. elegans. Science *298*, 830–834.

Dimitrov, A., Paupe, V., Gueudry, C., Sibarita, J.-B., Raposo, G., Vielemeyer, O., Gilbert, T., Csaba, Z., Attie-Bitach, T., Cormier-Daire, V., et al. (2009). The gene responsible for Dyggve-Melchior-Clausen syndrome encodes a novel peripheral membrane protein dynamically associated with the Golgi apparatus. Hum. Mol. Genet. *18*, 440–453.

Dornan, D., Wertz, I., Shimizu, H., Arnott, D., Frantz, G.D., Dowd, P., O'Rourke, K., Koeppen, H., and Dixit, V.M. (2004). The ubiquitin ligase COP1 is a critical negative regulator of p53. Nature *429*, 86–92.

Dupuis, N., Fafouri, A., Bayot, A., Kumar, M., Lecharpentier, T., Ball, G., Edwards, D., Bernard, V., Dournaud, P., Drunat, S., et al. (2015). Dymeclin deficiency causes postnatal microcephaly, hypomyelination and reticulum-to-Golgi trafficking defects in mice and humans. Hum. Mol. Genet.

Dyggve, H. V, Melchior, J.C., and Clausen, J. (1962). Morquio-Ullrich's Disease: An Inborn Error of Metabolism? Arch. Dis. Child. *37*, 525–534.

Eckler, A.M., Wilder, C., Castanon, A., Ferris, V.M., Lamere, R.A., Perrin, B.A., Pearlman, R., White, B., Byrd, C., Ludvik, N., et al. (2013). Haploinsufficiency of the Sec7 guanine nucleotide exchange factor gea1 impairs septation in fission yeast. PloS One *8*, e56807.

El Ghouzzi, V., Dagoneau, N., Kinning, E., Thauvin-Robinet, C., Chemaitilly, W., Prost-Squarcioni, C., Al-Gazali, L.I., Verloes, A., Le Merrer, M., Munnich, A., et al. (2003). Mutations in a novel gene Dymeclin (FLJ20071) are responsible for Dyggve–Melchior–Clausen syndrome. Hum. Mol. Genet. *12*, 357–364.

Emr, S., Glick, B.S., Linstedt, A.D., Lippincott-Schwartz, J., Luini, A., Malhotra, V., Marsh, B.J., Nakano, A., Pfeffer, S.R., Rabouille, C., et al. (2009). Journeys through the Golgi—taking stock in a new era. J. Cell Biol. *187*, 449–453.

Farquhar, M.G., and Palade, G.E. (1981). The Golgi apparatus (complex)-(1954-1981)-from artifact to center stage. J. Cell Biol. *91*, 77s – 103s.

Feinstein, M., Flusser, H., Lerman-Sagie, T., Ben-Zeev, B., Lev, D., Agamy, O., Cohen, I., Kadir, R., Sivan, S., Leshinsky-Silver, E., et al. (2014). VPS53 mutations cause progressive cerebello-cerebral atrophy type 2 (PCCA2). J. Med. Genet. *51*, 303–308.

Ferlay, J., Soerjomataram I, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman D, D., and Bray, F. (2014). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer J. Int. Cancer *136*, E359–E386.

Fielenbach, N., and Antebi, A. (2008). C. elegans dauer formation and the molecular basis of plasticity. Genes Dev. *22*, 2149–2165.

Foster, L.J., de Hoog, C.L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V.K., and Mann, M. (2006). A mammalian organelle map by protein correlation profiling. Cell *125*, 187–199.

Foulquier, F., Amyere, M., Jaeken, J., Zeevaert, R., Schollen, E., Race, V., Bammens, R., Morelle, W., Rosnoblet, C., Legrand, D., et al. (2012). TMEM165 deficiency causes a congenital disorder of glycosylation. Am. J. Hum. Genet. *91*, 15–26.

Fuchs, G., Diges, C., Kohlstaedt, L.A., Wehner, K.A., and Sarnow, P. (2011). Proteomic analysis of ribosomes: translational control of mRNA populations by glycogen synthase GYS1. J. Mol. Biol. *410*, 118–130.

Fukino, K., Iido, A., Teramoto, A., Sakamoto, G., Kasumi, F., Nakamura, Y., and Emi, M. (1999). Frequent allelic loss at the TOC locus on 17q25.1 in primary breast cancers. Genes. Chromosomes Cancer *24*, 345–350.

Fukui, Y., Kozasa, T., Kaziro, Y., Takeda, T., and Yamamoto, M. (1986). Role of a ras homolog in the life cycle of Schizosaccharomyces pombe. Cell *44*, 329–336.

Gabrielli, N., Ayté, J., and Hidalgo, E. (2012). Cells lacking pfh1, a fission yeast homolog of mammalian frataxin protein, display constitutive activation of the iron starvation response. J. Biol. Chem. *287*, 43042–43051.

García, P., Tajadura, V., García, I., and Sánchez, Y. (2006). Role of Rho GTPases and Rho-GEFs in the regulation of cell shape and integrity in fission yeast. Yeast Chichester Engl. *23*, 1031–1043.

Garcia, P., Tajadura, V., and Sanchez, Y. (2009). The Rho1p exchange factor Rgf1p signals upstream from the Pmk1 mitogen-activated protein kinase pathway in fission yeast. Mol. Biol. Cell *20*, 721–731.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387–391.

Glick, B.S. (2000). Organization of the Golgi apparatus. Curr. Opin. Cell Biol. *12*, 450–456.

Glick, B.S., and Luini, A. (2011). Models for Golgi Traffic: A Critical Assessment. Cold Spring Harb. Perspect. Biol. *3*.

Gonatas, N.K., Stieber, A., and Gonatas, J.O. (2006). Fragmentation of the Golgi apparatus in neurodegenerative diseases and cell death. J. Neurol. Sci. *246*, 21–30.

Gray, J. V, Petsko, G.A., Johnston, G.C., Ringe, D., Singer, R.A., and Werner-Washburne, M. (2004). “Sleeping beauty”: quiescence in Saccharomyces cerevisiae. Microbiol. Mol. Biol. Rev. MMBR *68*, 187–206.

Griffin, J.L. (2004). Metabolic profiles to define the genome : can we hear the phenotypes ?

Griffiths-Jones, S. (2007). Annotating noncoding RNA genes. Annu. Rev. Genomics Hum. Genet. *8*, 279–298.

Grosset, C., Chen, C.Y., Xu, N., Sonenberg, N., Jacquemin-Sablon, H., and Shyu, A.B. (2000). A mechanism for translationally coupled mRNA turnover: interaction between the poly(A) tail and a c-fos RNA coding determinant via a protein complex. Cell *103*, 29–40.

Gutteridge, A., Pir, P., Castrillo, J.I., Charles, P.D., Lilley, K.S., and Oliver, S.G. (2010). Nutrient control of eukaryote cell growth: a systems biology study in yeast. BMC Biol. *8*, 68.

Gygi, S.P. (2000). Using mass spectrometry for quantitative proteomics. Trends Guide *1931*, 31–36.

Harada, H., Nagai, H., and Tsuneizumi, M. (2001). Identification of DMC1 , a novel gene in the TOC region on 17q25 . 1 that shows loss of expression in multiple human cancers. J. Hum. Genet. *46*, 90–95.

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. Annu. Rev. Biochem. *67*, 425–479.

Hoffmann, K., Pawłowska, J., Walther, G., Wrzosek, M., de Hoog, G.S., Benny, G.L., Kirk, P.M., and Voigt, K. (2013). The family structure of the Mucorales: a synoptic revision based on comprehensive multigene-genealogies. Persoonia *30*, 57–76.

Hollstein, M., Rice, K., Greenblatt, M.S., Soussi, T., Fuchs, R., Sørlie, T., Hovig, E., Smith-Sørensen, B., Montesano, R., and Harris, C.C. (1994). Database of p53 gene somatic mutations in human tumors and cell lines. Nucleic Acids Res. *22*, 3551–3555.

Horák, J. (2003). The role of ubiquitin in down-regulation and intracellular sorting of membrane proteins: insights from yeast. Biochim. Biophys. Acta *1614*, 139–155.

van der Horst, A., de Vries-Smits, A.M.M., Brenkman, A.B., van Triest, M.H., van den Broek, N., Colland, F., Maurice, M.M., and Burgering, B.M.T. (2006). FOXO4 transcriptional activity is regulated by monoubiquitination and USP7/HAUSP. Nat. Cell Biol. *8*, 1064–1073.

Hughes, A.L., Todd, B.L., and Espenshade, P.J. (2005). SREBP pathway responds to sterols and functions as an oxygen sensor in fission yeast. Cell *120*, 831–842.

Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. *15*, 205–213.

Inoue, K. (2005). PLP1-related inherited dysmyelinating disorders: Pelizaeus-Merzbacher disease and spastic paraplegia type 2. Neurogenetics *6*, 1–16.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., et al. (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. *37*, D412–D416.

Jones, R.G., and Thompson, C.B. (2009). Tumor suppressors and cell metabolism: a recipe for cancer growth. Genes Dev. *23*, 537–548.

Joshi, G., Chi, Y., Huang, Z., and Wang, Y. (2014). Aβ-induced Golgi fragmentation in Alzheimer's disease enhances Aβ production. Proc. Natl. Acad. Sci. U. S. A. *111*, E1230–E1239.

Jura, N., Scotto-Lavino, E., Sobczyk, A., and Bar-Sagi, D. (2006). Differential modification of Ras proteins by ubiquitination. Mol. Cell *21*, 679–687.

Kadura, S., He, X., Vanoosthuyse, V., Hardwick, K.G., and Sazer, S. (2005). The A78V mutation in the Mad3-like domain of Schizosaccharomyces pombe Bub1p perturbs nuclear accumulation and kinetochore targeting of Bub1p, Bub3p, and Mad3p and spindle assembly checkpoint function. Mol. Biol. Cell *16*, 385–395.

Kaletta, T., and Hengartner, M.O. (2006). Finding function in novel targets: C. elegans as a model organism. Nat. Rev. Drug Discov. *5*, 387–398.

Kandziora, F., Neumann, L., Schnake, K.J., Khodadadyan-Klostermann, C., Rehart, S., Haas, N.P., and Mittlmeier, T. (2002). Atlantoaxial instability in Dyggve-Melchior-Clausen syndrome. Case report and review of the literature. J. Neurosurg. *96*, 112–117.

Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. *9*, 605–618.

Kell, D.B. (2004). Metabolomics and systems biology: making sense of the soup. Curr. Opin. Microbiol. *7*, 296–307.

Kelley, R., and Ideker, T. (2005). Systematic interpretation of genetic interactions using protein networks. Nat. Biotechnol. *23*, 561–566.

Kellner, M., Burmester, A., Wöstemeyer, A., and Wöstemeyer, J. (1993). Transfer of genetic information from the mycoparasite Parasitella parasitica to its host Absidia glauca. Curr. Genet. *23*, 334–337.

Kennedy, P.J., Vashisht, A.A., Hoe, K.-L., Kim, D.-U., Park, H.-O., Hayles, J., and Russell, P. (2008). A genome-wide screen of genes involved in cadmium tolerance in Schizosaccharomyces pombe. Toxicol. Sci. Off. J. Soc. Toxicol. *106*, 124–139.

Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. (2009). Human Protein Reference Database--2009 update. Nucleic Acids Res. *37*, D767–D772.

Khalifa, O., Imtiaz, F., Al-Sakati, N., Al-Manea, K., Verloes, A., and Al-Owain, M. (2011). Dyggve-Melchior-Clausen syndrome: novel splice mutation with atlanto-axial subluxation. Eur. J. Pediatr. *170*, 121–126.

Kim, E.K., and Choi, E.-J. (2010). Pathological roles of MAPK signaling pathways in human diseases. Biochim. Biophys. Acta *1802*, 396–405.

Kim, W., Bennett, E.J., Huttlin, E.L., Guo, A., Li, J., Possemato, A., Sowa, M.E., Rad, R., Rush, J., Comb, M.J., et al. (2011). Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol. Cell *44*, 325–340.

King, M.-C. (2014). “The race” to clone BRCA1. Science *343*, 1462–1465.

Klaunig, J.E., Kamendulis, L.M., and Hocevar, B.A. (2010). Oxidative stress and oxidative damage in carcinogenesis. Toxicol. Pathol. *38*, 96–109.

Kleene, R., and Berger, E.G. (1993). The molecular and cell biology of glycosyltransferases. Biochim. Biophys. Acta *1154*, 283–325.

Klute, M.J., Melançon, P., and Dacks, J.B. (2011). Evolution and Diversity of the Golgi. Cold Spring Harb. Perspect. Biol. *3*, a007849.

Kobayashi, H., Kawauchi, D., Hashimoto, Y., Ogata, T., and Murakami, F. (2013). The control of precerebellar neuron migration by RNA-binding protein Csde1. Neuroscience *253*, 292–303.

Komander, D., Clague, M.J., and Urbé, S. (2009). Breaking the chains: structure and function of the deubiquitinases. Nat. Rev. Mol. Cell Biol. *10*, 550–563.

Koonin, E. V (2010). The origin and early evolution of eukaryotes in the light of phylogenomics. Genome Biol. *11*, 209.

Kornitzer, D., and Ciechanover, A. (2000). Modes of regulation of ubiquitin-mediated protein degradation. J. Cell. Physiol. *182*, 1–11.

Koul, H.K., Pal, M., and Koul, S. (2013). Role of p38 MAP Kinase Signal Transduction in Solid Tumors. Genes Cancer *4*, 342–359.

Kouranti, I., McLean, J.R., Feoktistova, A., Liang, P., Johnson, A.E., Roberts-Galbraith, R.H., and Gould, K.L. (2010). A global census of fission yeast deubiquitinating enzyme localization and interaction networks reveals distinct compartmentalization profiles and overlapping functions in endocytosis and polarity. PLoS Biol. *8*.

Kushi, L.H., Byers, T., Doyle, C., Bandera, E. V., McCullough, M., Gansler, T., Andrews, K.S., and Thun, M.J. (2006). American Cancer Society Guidelines on Nutrition and Physical Activity for Cancer Prevention: Reducing the Risk of Cancer With Healthy Food Choices and Physical Activity. CA. Cancer J. Clin. *56*, 254–281.

Kussmann, M., Rezzi, S., and Daniel, H. (2008). Profiling techniques in nutrition and health research. Curr Opin Biotechnol *19*, 83–99.

Lackner, D.H., Schmidt, M.W., Wu, S., Wolf, D.A., and Bähler, J. (2012). Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast. Genome Biol. *13*, R25.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

Lambou, K., Pennati, A., Valsecchi, I., Tada, R., Sherman, S., Sato, H., Beau, R., Gadda, G., and Latge, J.-P. (2013). Pathway of Glycine Betaine Biosynthesis in Aspergillus fumigatus. Eukaryot. Cell *12*, 853–863.

Langella, O., Valot, B., Jacob, D., Balliau, T., Flores, R., Hoogland, C., Joets, J., and Zivy, M. (2013). Management and dissemination of MS proteomic data with PROTICdb: example of a quantitative comparison between methods of protein extraction. Proteomics *13*, 1457–1466.

Lassout, O., Hochstrasser, D., and Lescuyer, P. (2014). Assessment of the influence of the patient's inflammatory state on the accuracy of a haptoglobin selected reaction monitoring assay. Clin. Proteomics *11*.

Lee, E.Y.H.P., and Muller, W.J. (2010). Oncogenes and tumor suppressor genes. Cold Spring Harb. Perspect. Biol. *2*, a003236.

Leong, H.S., Dawson, K., Wirth, C., Li, Y., Connolly, Y., Smith, D.L., Wilkinson, C.R.M., and Miller, C.J. (2014). A global non-coding RNA system modulates fission yeast protein levels in response to stress. Nat. Commun. *5*, 3947.

Li, M., Chen, D., Shiloh, A., Luo, J., Nikolaev, A.Y., Qin, J., and Gu, W. (2002). Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. Nature *416*, 648–653.

Li, W., Bengtson, M.H., Ulbrich, A., Matsuda, A., Reddy, V.A., Orth, A., Chanda, S.K., Batalov, S., and Joazeiro, C.A.P. (2008a). Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. PloS One *3*, e1487.

Li, Y., Guessous, F., Kwon, S., Kumar, M., Ibidapo, O., Fuller, L., Johnson, E., Lal, B., Hussaini, I., Bao, Y., et al. (2008b). PTEN has tumor-promoting properties in the setting of gain-of-function p53 mutations. Cancer Res. *68*, 1723–1731.

Liegel, R.P., Handley, M.T., Ronchetti, A., Brown, S., Langemeyer, L., Linford, A., Chang, B., Morris-Rosendahl, D.J., Carpanini, S., Posmyk, R., et al. (2013). Loss-of-function mutations in TBC1D20 cause cataracts and male infertility in blind sterile mice and Warburg micro syndrome in humans. Am. J. Hum. Genet. *93*, 1001–1014.

Liu, S., and Storrie, B. (2012). Are Rab proteins the link between Golgi organization and membrane trafficking? Cell. Mol. Life Sci. CMLS *69*, 4093–4106.

Losev, E., Reinke, C.A., Jellen, J., Strongin, D.E., Bevis, B.J., and Glick, B.S. (2006). Golgi maturation visualized in living yeast. Nature *441*, 1002–1006.

Maglott, D., Ostell, J., Pruitt, K.D., and Tatusova, T. (2011). Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. *39*, D52–D57.

Malumbres, M., and Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. Nat. Rev. Cancer *9*, 153–166.

Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., and Bähler, J. (2012). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell *151*, 671–683.

Martínez-Alonso, E., Tomás, M., and Martínez-Menárguez, J.A. (2013). Morpho-functional architecture of the Golgi complex of neuroendocrine cells. Front. Endocrinol. *4*, 41.

Mata, J., Marguerat, S., and Bähler, J. (2005). Post-transcriptional control of gene expression: a genome-wide perspective. Trends Biochem. Sci. *30*, 506–514.

Matsuura-Tokita, K., Takeuchi, M., Ichihara, A., Mikuriya, K., and Nakano, A. (2006). Live imaging of yeast Golgi cisternal maturation. Nature *441*, 1007–1010.

Matsuyama, A., Arai, R., Yashiroda, Y., Shirai, A., Kamata, A., Sekido, S., Kobayashi, Y., Hashimoto, A., Hamamoto, M., Hiraoka, Y., et al. (2006). ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. *24*, 841–847.

McGrath, P.T., Xu, Y., Ailion, M., Garrison, J.L., Butcher, R.A., and Bargmann, C.I. (2011). Parallel evolution of domesticated Caenorhabditis species targets pheromone receptor genes. Nature *477*, 321–325.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. *41*, W597–W600.

Mesa, R., Luo, S., Hoover, C.M., Miller, K., Minniti, A., Inestrosa, N., and Nonet, M.L. (2011). HID-1, a new component of the peptidergic signaling pathway. Genetics *187*, 467–483.

Michaillat, L., and Mayer, A. (2013). Identification of genes affecting vacuole membrane fragmentation in Saccharomyces cerevisiae. PloS One *8*, e54160.

Mihailovich, M., Militti, C., Gabaldón, T., and Gebauer, F. (2010). Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. BioEssays News Rev. Mol. Cell. Dev. Biol. *32*, 109–118.

Momand, J., Zambetti, G.P., Olson, D.C., George, D., and Levine, A.J. (1992). The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. Cell *69*, 1237–1245.

Morigasaki, S., Shimada, K., Ikner, A., Yanagida, M., and Shiozaki, K. (2008). Glycolytic enzyme GAPDH promotes peroxide stress signaling through multistep phosphorelay to a MAPK cascade. Mol. Cell *30*, 108–113.

Nakae, J., Oki, M., and Cao, Y. (2008). The FoxO transcription factors and metabolic regulation. FEBS Lett. *582*, 54–67.

Nakano, A., and Luini, A. (2010). Passage through the Golgi. Curr. Opin. Cell Biol. *22*, 471–478.

Naresh, A., Saini, S., and Singh, J. (2003). Identification of Uhp1, a ubiquitinated histone-like protein, as a target/mediator of Rhp6 in mating-type silencing in fission yeast. J. Biol. Chem. *278*, 9185–9194.

Nguyen, A.N., Ikner, A.D., Shiozaki, M., Warren, S.M., and Shiozaki, K. (2002). Cytoplasmic localization of Wis1 MAPKK by nuclear export signal is important for nuclear targeting of Spc1/Sty1 MAPK in fission yeast. Mol. Biol. Cell *13*, 2651–2663.

Nielsen, O., Davey, J., and Egel, R. (1992). The ras1 function of Schizosaccharomyces pombe mediates pheromone-induced transcription. EMBO J. *11*, 1391–1395.

Nilsson, T., Au, C.E., and Bergeron, J.J.M. (2009). Sorting out glycosylation enzymes in the Golgi apparatus. FEBS Lett. *583*, 3764–3769.

Novak, B., Csikasz-Nagy, A., Gyorffy, B., Chen, K., and Tyson, J.J. (1998). Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions. Biophys. Chem. *72*, 185–200.

Ohmiya, R., Yamada, H., Nakashima, K., Aiba, H., and Mizuno, T. (1995). Osmoregulation of fission yeast: cloning of two distinct genes encoding glycerol-3-phosphate dehydrogenase, one of which is responsible for osmotolerance for growth. Mol. Microbiol. *18*, 963–973.

Onken, B., Wiener, H., Philips, M.R., and Chang, E.C. (2006). Compartmentalized signaling of Ras in fission yeast. Proc. Natl. Acad. Sci. U. S. A. *103*, 9045–9050.

Osipovich, A.B., Jennings, J.L., Lin, Q., Link, A.J., and Ruley, H.E. (2008). Dyggve-Melchior-Clausen syndrome: chondrodysplasia resulting from defects in intracellular vesicle traffic. Proc. Natl. Acad. Sci. U. S. A. *105*, 16171–16176.

Pandey, A., and Mann, M. (2000). Proteomics to study genes and genomes. Nature *405*, 837–846.

Parr, E. (2012). The default state of the cell: quiescence or proliferation? BioEssays News Rev. Mol. Cell. Dev. Biol. *34*, 36–37.

Patterson, G.I., and Padgett, R.W. (2000). TGF beta-related pathways. Roles in Caenorhabditis elegans development. Trends Genet. TIG *16*, 27–33.

Patterson, S.D., and Aebersold, R.H. (2003). Proteomics: the first decade and beyond. Nat. Genet. *33 Suppl*, 311–323.

Perdacher, M. (2011). Next-Generation Sequencing and its Applications in RNA-Seq. Aboutme.biobyte.org.

Pilot-Storck, F., Chopin, E., Rual, J.-F., Baudot, A., Dobrokhotov, P., Robinson-Rechavi, M., Brun, C., Cusick, M.E., Hill, D.E., Schaeffer, L., et al. (2010). Interactome mapping of the

phosphatidylinositol 3-kinase-mammalian target of rapamycin pathway identifies deformed epidermal autoregulatory factor-1 as a new glycogen synthase kinase-3 interactor. Mol. Cell. Proteomics MCP *9*, 1578–1593.

Pluskal, T., Nakamura, T., Villar-Briones, A., and Yanagida, M. (2010). Metabolic profiling of the fission yeast S. pombe: quantification of compounds under different temperatures and genetic perturbation. Mol. Biosyst. *6*, 182–198.

Preuss, D., Mulholland, J., Franzusoff, A., Segev, N., and Botstein, D. (1992). Characterization of the Saccharomyces Golgi complex through the cell cycle by immunoelectron microscopy. Mol. Biol. Cell *3*, 789–803.Rabouille, C., Hui, N., Hunte, F., Kieckbusch, R., Berger, E.G., Warren, G., and Nilsson, T. (1995a). Mapping the distribution of Golgi enzymes involved in the construction of complex oligosaccharides. J. Cell Sci. *108 (Pt 4)*, 1617–1627.

Risk, J.M., Ruhrberg, C., Hennies, H., Mills, H.S., Di Colandrea, T., Evans, K.E., Ellis, A., Watt, F.M., Bishop, D.T., Spurr, N.K., et al. (1999). Envoplakin, a possible candidate gene for focal NEPPK/esophageal cancer (TOC): the integration of genetic and physical maps of the TOC region on 17q25. Genomics *59*, 234–242.

Robinson, N.G., Guo, L., Imai, J., Toh-E, A., Matsui, Y., and Tamanoi, F. (1999). Rho3 of Saccharomyces cerevisiae, which regulates the actin cytoskeleton and exocytosis, is a GTPase which interacts with Myo2 and Exo70. Mol. Cell. Biol. *19*, 3580–3587.

Rogowska-Wrzesinska, A., Le Bihan, M.-C., Thaysen-Andersen, M., and Roepstorff, P. (2013). 2D gels still have a niche in proteomics. J. Proteomics *88*, 4–13.

Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S.R., Qu, H., Shales, M., Park, H.-O., Hayles, J., et al. (2008). Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. Science *322*, 405–410.

Rossanese, O.W., Reinke, C.A., Bevis, B.J., Hammond, A.T., Sears, I.B., O'Connor, J., and Glick, B.S. (2001). A role for actin, Cdc1p, and Myo2p in the inheritance of late Golgi elements in Saccharomyces cerevisiae. J. Cell Biol. *153*, 47–62.

Rustici, G., Mata, J., Kivinen, K., Lió, P., Penkett, C.J., Burns, G., Hayles, J., Brazma, A., Nurse, P., and Bähler, J. (2004). Periodic gene expression program of the fission yeast cell cycle. Nat. Genet. *36*, 809–817.

Ryan, C.J., Roguev, A., Patrick, K., Xu, J., Jahari, H., Tong, Z., Beltrao, P., Shales, M., Qu, H., Collins, S.R., et al. (2012). Hierarchical modularity and the evolution of genetic interactomes across species. Mol. Cell *46*, 691–704.

Sáenz, J.B., Sun, W.J., Chang, J.W., Li, J., Bursulaya, B., Gray, N.S., and Haslam, D.B. (2009). Golgicide A reveals essential roles for GBF1 in Golgi assembly and function. Nat. Chem. Biol. *5*, 157–165.

Sajiki, K., Hatanaka, M., Nakamura, T., Takeda, K., Shimanuki, M., Yoshida, T., Hanyu, Y., Hayashi, T., Nakaseko, Y., and Yanagida, M. (2009). Genetic control of cellular quiescence in S. pombe. J. Cell Sci. *122*, 1418–1429.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: A laboratory manual+ Cold Spring Harbor (New York: Cold Spring Harbor Laboratory Press).

Santos, H.G., Fernandes, H.C., Nunes, J.L., and Almeida, M.R. (2009). Portuguese case of Smith-McCort syndrome caused by a new mutation in the Dymeclin (FLJ20071) gene. Clin. Dysmorphol. *18*, 41–44.

Schmidt, M.W., Houseman, A., Ivanov, A.R., and Wolf, D.A. (2007). Comparative proteomic and transcriptomic profiling of the fission yeast Schizosaccharomyces pombe. Mol. Syst. Biol. *3*, 79.

Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., et al. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell *123*, 507–519.

Schulze, W.X., and Usadel, B. (2010). Quantitation in mass-spectrometry-based proteomics. Annu. Rev. Plant Biol. *61*, 491–516.

Seemann, J., Jokitalo, E., Pypaert, M., and Warren, G. (2000). Matrix proteins can generate the higher order architecture of the Golgi apparatus. Nature *407*, 1022–1026.

Servant, F., Bru, C., Carrère, S., Courcelle, E., Gouzy, J., Peyruc, D., and Kahn, D. (2002). ProDom: automated clustering of homologous domains. Brief. Bioinform. *3*, 246–251.

Shaid, S., Brandts, C.H., Serve, H., and Dikic, I. (2013). Ubiquitination and selective autophagy. Cell Death Differ. *20*, 21–30.

Shaw, R.J., and Cantley, L.C. (2006). Ras, PI(3)K and mTOR signalling controls tumour cell growth. Nature *441*, 424–430.

Sherr, C.J. (2004). Principles of tumor suppression. Cell *116*, 235–246.

Shiloh, Y. (2003). ATM and related protein kinases: safeguarding genome integrity. Nat. Rev. Cancer *3*, 155–168.

Shiozaki, K., and Russell, P. (1995). Cell-cycle control linked to extracellular environment by MAP kinase pathway in fission yeast. Nature *378*, 739–743.

Sideri, T., Rallis, C., Bitton, D.A., Lages, B.M., Suo, F., Rodríguez-López, M., Du, L.-L., and Bähler, J. (2014). Parallel profiling of fission yeast deletion mutants for proliferation and for lifespan during long-term quiescence. G3 Bethesda Md *5*, 145–155.

Soetens, O., De Craene, J.O., and Andre, B. (2001). Ubiquitin is required for sorting to the vacuole of the yeast general amino acid permease, Gap1. J. Biol. Chem. *276*, 43949–43957.

Søgaard, C., Stenbæk, A., Bernard, S., Hadi, M., Driouich, A., Scheller, H.V., and Sakuragi, Y. (2012). GO-PROMTO illuminates protein membrane topologies of glycan biosynthetic enzymes in the Golgi apparatus of living tissues. PloS One *7*, e31324.

Song, M.S., Salmena, L., Carracedo, A., Egia, A., Lo-Coco, F., Teruya-Feldstein, J., and Pandolfi, P.P. (2008). The deubiquitinylation and localization of PTEN are regulated by a HAUSP-PML network. Nature *455*, 813–817.

Staehelin, L.A., and Kang, B.-H. (2008). Nanoscale architecture of endoplasmic reticulum export sites and of Golgi membranes as determined by electron tomography. Plant Physiol. *147*, 1454–1468.

Stewart, E.V., Nwosu, C.C., Tong, Z., Roguev, A., Cummins, T.D., Kim, D.-U., Hayles, J., Park, H.-O., Hoe, K.-L., Powell, D.W., et al. (2011). Yeast SREBP cleavage activation requires the Golgi Dsc E3 ligase complex. Mol. Cell *42*, 160–171.

Strathmann, F.G., and Hoofnagle, A.N. (2011). Current and Future Applications of Mass Spectrometry to the Clinical Laboratory. Am. J. Clin. Pathol. *136*, 609–616.

Su, L., Liu, X., Chai, N., Lv, L., Wang, R., Li, X., Nie, Y., Shi, Y., and Fan, D. (2014). The transcription factor FOXO4 is down-regulated and inhibits tumor proliferation and metastasis in gastric cancer. BMC Cancer *14*, 378.

Suda, Y., and Nakano, A. (2012). The yeast Golgi apparatus. Traffic Cph. Den. *13*, 505–510.

Sun, K.-H., de Pablo, Y., Vincent, F., Johnson, E.O., Chavers, A.K., and Shah, K. (2008). Novel genetic tools reveal Cdk5's major role in Golgi fragmentation in Alzheimer's disease. Mol. Biol. Cell *19*, 3052–3069.

Sun, L.-L., Li, M., Suo, F., Liu, X.-M., Shen, E.-Z., Yang, B., Dong, M.-Q., He, W.-Z., and Du, L.-L. (2013). Global analysis of fission yeast mating genes reveals new autophagy factors. PLoS Genet. *9*, e1003715.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS ONE *6*.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Takeda, K., Mori, A., and Yanagida, M. (2011). Identification of genes affecting the toxicity of anti-cancer drug bortezomib by genome-wide screening in S. pombe. PloS One *6*, e22021.

Tamanoi, F. (2011). Ras signaling in yeast. Genes Cancer *2*, 210–215.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. *30*, 2725–2729.

Ting, C.-H., Wen, H.-L., Liu, H.-C., Hsieh-Li, H.-M., Li, H., and Lin-Chao, S. (2012). The spinal muscular atrophy disease protein SMN is linked to the Golgi network. PloS One *7*, e51826.

Todd, B.L., Stewart, E. V., Burg, J.S., Hughes, A.L., and Espenshade, P.J. (2006). Sterol regulatory element binding protein is a principal regulator of anaerobic gene expression in fission yeast. Mol. Cell. Biol. *26*, 2817–2831.

Tu, Y., Chen, C., Pan, J., Xu, J., Zhou, Z.-G., and Wang, C.-Y. (2012). The Ubiquitin Proteasome Pathway (UPP) in the regulation of cell cycle control and DNA damage repair and its implication in tumorigenesis. Int. J. Clin. Exp. Pathol. *5*, 726–738.

Ungar, D. (2009). Golgi linked protein glycosylation and associated diseases. Semin. Cell Dev. Biol. *20*, 762–769.

Valot, B., Langella, O., Nano, E., and Zivy, M. (2011). MassChroQ: a versatile tool for mass spectrometry quantification. Proteomics *11*, 3572–3577.

Vinke, F.P., Grieve, A.G., and Rabouille, C. (2011). The multiple facets of the Golgi reassembly stacking proteins. Biochem. J. *433*, 423–433.

Wang, L., Zhan, Y., Song, E., Yu, Y., Jiu, Y., Du, W., Lu, J., Liu, P., Xu, P., and Xu, T. (2011). HID-1 is a peripheral membrane protein primarily associated with the medial- and trans- Golgi apparatus. Protein Cell *2*, 74–85.

Washburn, M.P. (2011). Driving biochemical discovery with quantitative proteomics. Trends Biochem. Sci. *36*, 170–177.

Watanabe, M., Watanabe, D., Nogami, S., Morishita, S., and Ohya, Y. (2009). Comprehensive and quantitative analysis of yeast deletion mutants defective in apical and isotropic bud growth. Curr. Genet. *55*, 365–380.

Weckwerth, W., Wenzel, K., and Fiehn, O. (2004). Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. Proteomics *4*, 78–83.

Weeks, M.E., Sinclair, J., Butt, A., Chung, Y.-L., Worthington, J.L., Wilkinson, C.R.M., Griffiths, J., Jones, N., Waterfield, M.D., and Timms, J.F. (2006). A parallel proteomic and metabolomic analysis of the hydrogen peroxide- and Sty1p-dependent stress response in Schizosaccharomyces pombe. Proteomics *6*, 2772–2796.

Weissman, A.M. (2001). Themes and variations on ubiquitylation. Nat. Rev. Mol. Cell Biol. *2*, 169–178.

Wilkins, A.S., and Holliday, R. (2009). The Evolution of Meiosis From Mitosis. Genetics *181*, 3–12.

Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., and Williams, K.L. (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol. Genet. Eng. Rev. *13*, 19–50.

Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., et al. (2002). The genome sequence of Schizosaccharomyces pombe. Nature *415*, 871–880.

Wood, V., Harris, M.A., McDowall, M.D., Rutherford, K., Vaughan, B.W., Staines, D.M., Aslett, M., Lock, A., Bähler, J., Kersey, P.J., et al. (2012). PomBase: a comprehensive online resource for fission yeast. Nucleic Acids Res. *40*, D695–D699.

Wooley, J.C., Lin, H.S., and Biology, N.R.C. (US) C. on F. at the I. of C. and (2005). Executive Summary.

Xu, G., and Jaffrey, S.R. (2011). The new landscape of protein ubiquitination. Nat. Biotechnol. *29*, 1098–1100.

Yanagida, M. (2002). The model unicellular eukaryote, Schizosaccharomyces pombe. Genome Biol. *3*, COMMENT2003.

Yoon, H.S., Grant, J., Tekle, Y.I., Wu, M., Chaon, B.C., Cole, J.C., Logsdon, J.M., Patterson, D.J., Bhattacharya, D., and Katz, L.A. (2008). Broadly sampled multigene trees of eukaryotes. BMC Evol. Biol. *8*, 14.

Yu, Y., Wang, L., Jiu, Y., Zhan, Y., Liu, L., Xia, Z., Song, E., Xu, P., and Xu, T. (2011). HID-1 is a novel player in the regulation of neuropeptide sorting. Biochem. J. *434*, 383–390.

Zaitsevskaya-Carter, T., and Cooper, J.A. (1997). Spm1, a stress-activated MAP kinase that regulates morphogenesis in S.pombe. EMBO J. *16*, 1318–1331.

Zhang, C., Li, D., Zhang, J., Chen, X., Huang, M., Archacki, S., Tian, Y., Ren, W., Mei, A., Zhang, Q., et al. (2013a). Mutations in ABCB6 cause dyschromatosis universalis hereditaria. J. Invest. Dermatol. *133*, 2221–2228.

Zhang, Z.-R., Bonifacino, J.S., and Hegde, R.S. (2013b). Deubiquitinases sharpen substrate discrimination during membrane protein degradation from the ER. Cell *154*, 609–622.

Zhao, L.Y., Liu, J., Sidhu, G.S., Niu, Y., Liu, Y., Wang, R., and Liao, D. (2004). Negative regulation of p53 functions by Daxx and the involvement of MDM2. J. Biol. Chem. *279*, 50566–50579.

Zhou, X., Ma, Y., Sugiura, R., Kobayashi, D., Suzuki, M., Deng, L., and Kuno, T. (2010). MAP kinase kinase kinase (MAPKKK)-dependent and -independent activation of Sty1 stress MAPK in fission yeast. J. Biol. Chem. *285*, 32818–32823.